BEST AVAILABLE COPY

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:39:57 ; Search time 21 Seconds
(without alignments)
2238.559 Million cell updates/sec

Title: US-09-455-978B-2
Perfect score: 2394
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVRTVEEVRETVGKLS 489

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2394	100.0	489	2	T44978	transducer protein
2	2360	98.6	489	2	E84304	Htr10 transducer [
3	643	26.9	481	2	A84294	Htr9 transducer [i
4	609.5	25.5	482	2	T44973	transducer protein
5	597	24.9	633	2	H84336	Htr3 transducer [i
6	597	24.9	805	2	T48840	transducer protein
7	580	24.2	790	2	B84238	Htr18 transducer [
8	578	24.1	778	2	F84237	Htr4 transducer [i
9	573	23.9	778	2	T48897	transducer protein
10	560.5	23.4	452	2	T44849	transducer protein
11	555	23.2	810	2	F84327	Htr5 transducer [i
12	554	23.1	810	2	T46810	halobacterial tran
13	540.5	22.6	777	2	T44597	transducer protein
14	537.5	22.5	534	2	S55299	sensory rhodopsin
15	535.5	22.4	636	2	A84252	Htr15 transducer [
16	529	22.1	804	2	T44606	transducer protein
17	524	21.9	451	2	T44964	transducer protein
18	523.5	21.9	544	2	T44938	transducer protein
19	523.5	21.9	643	2	H84305	Htr8 transducer [i
20	521.5	21.8	628	2	F84219	Htr16 transducer [
21	521	21.8	545	2	E84327	Htr7 transducer [i
22	521	21.8	545	2	T46811	halobacterial tran
23	509.5	21.3	789	2	E84236	Htr6 transducer [i
24	507.5	21.2	788	2	T44262	transducer protein
25	506	21.1	420	2	C84298	Htr12 transducer [
26	505	21.1	642	1	T44253	transducer protein
27	502.5	21.0	773	2	T44989	transducer protein
28	499.5	20.9	627	2	F84194	Htr14 transducer [
29	497.5	20.8	419	2	T44276	transducer protein
30	497	20.8	536	2	D84325	Htr17 transducer [
31	475.5	19.9	537	2	C87302	methyl-accepting c
32	464	19.4	764	2	A84328	Htr2 transducer [i
33	463.5	19.4	536	1	A47190	transducer protein
34	463.5	19.4	536	2	E84318	Htr1 transducer [i
35	461	19.3	765	1	T44946	transducer protein
36	452.5	18.9	423	2	H84257	Htr13 transducer [
37	439.5	18.4	423	2	T44258	transducer protein
38	419.5	17.5	559	2	G84132	methyl-accepting c
39	415	17.3	555	2	D87536	methyl-accepting c
40	413	17.3	439	2	A83713	methyl-accepting c
41	412.5	17.2	432	2	C69832	methyl-accepting c
42	409.5	17.1	632	2	H83106	chemotactic transd
43	408	17.0	499	2	A97485	methyl-accepting c
44	408	17.0	499	2	AB2703	methyl-accepting c
45	405.5	16.9	579	2	D84137	methyl-accepting c

ALIGNMENTS

RESULT 1
T44978
transducer protein hemAT [validated] - Halobacterium salinarum
N;Alternate names: methyl-accepting taxis protein htB; transducer protein htB; transd
C;Species: Halobacterium salinarum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 15-Sep-2000
C;Accession: T44978
R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A;Title: Signal transduction in the archaeon Halobacterium salinarium is processed th
A;Reference number: Z22804; MUID:96209786; PMID:8643458
A;Accession: T44978
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-489 <ZHA>
A;Cross-references: EMBL:U75436; NID:g1654420; PIDN:AAB17881.1; PID:g1654421
A;Experimental source: strain Flx15
A;Note: the source is designated as Halobacterium salinarium
C;Genetics:
A;Gene: hemAT; htp15; htB
C;Function:
A;Description: involved in aerotactic signal transduction; involved in oxygen sensing
C;Superfamily: Halobacterium salinarum transducer protein htrI
C;Keywords: heme; methylated amino acid; signal transduction

Query Match 100.0%; Score 2394; DB 2; Length 489;
Best Local Similarity 100.0%; Pred. No. 8.4e-98;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

Qy      61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120

Qy     121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180

Qy     181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRART 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRART 240

Qy     241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300

Qy     301 EATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 EATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360
```

```
Qy    361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420

Qy    421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEE 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEE 480

Qy    481 VRETVGKLS 489
          |||||||||
Db    481 VRETVGKLS 489


RESULT 2
E84304
Htr10 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: E84304
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: E84304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-489 <STO>
A;Cross-references: GB:AE004437; NID:g10580997; PIDN:AAG19801.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr10
C;Superfamily: Halobacterium salinarum transducer protein htrI

  Query Match            98.6%;  Score 2360;  DB 2;  Length 489;
  Best Local Similarity  99.0%;  Pred. No. 2.6e-96;
  Matches  484;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

Qy     61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120

Qy    121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180

Qy    181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRART 240
          ||||||||||||||||||||||||||||: |||||||||||||||||||||||||| | |
Db    181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVASDVEAPLSSLEATSQDVAERTDTMRAAT 240

Qy    241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300

Qy    301 EATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 EATDGVTAGVEQLRERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360

Qy    361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420

Qy    421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEE 480
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDDVDDRAGEIAAALDDIADATDQQVRTVEE 480

Qy    481 VRETVGKLS 489
          |||||||||
Db    481 VRETVGKLS 489


RESULT 3
A84294
Htr9 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
C;Accession: A84294
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: A84294
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-481 <STO>
A;Cross-references: GB:AE004437; NID:g10580901; PIDN:AAG19717.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr9
C;Superfamily: Halobacterium salinarum transducer protein car

  Query Match            26.9%;  Score 643;  DB 2;  Length 481;
  Best Local Similarity  37.6%;  Pred. No. 2.2e-21;
  Matches  182;  Conservative   72;  Mismatches  196;  Indels   34;  Gaps   10;

Qy     15 NGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADAL-VTDFYDHLE 73
          | :|  | |  :  | | :||        || :|||     |  |  ||: ||| : |
Db     19 NTLDVPAFA--VDADGAVVAW--------DDQIAAL-----LETAPEDAIGVTDIGERLN 63

Qy     74 SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVY 133
                | : |  :   :   ||    | |  |  ||            :    |:
Db     64 D----DGSRALANKVADTPIDAHHEY---DGVGLADESYALLTGDYVYEDTTVAGNTDLW 116

Qy    134 LGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLL------TFDQQI--- 184
            |  |:||    :  ::| ||   :|    || | |  ::  |       ||  :
Db    117 FIATPVYHTGEFRGVI-EIVQDR-SSSARYQSELQALFGELVDTLDAYDAGRFDATVDIA 174

Qy    185 AMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQV 244
          | || :|    ::   :    :   |  | |   :  ||| || |:| :  :   :  |
Db    175 AEDTLLDDEYIQIGRNLTEFGDTLAAHITEVHNDVERLEAASQAVSESSAEIDELSTAQS 234

Qy    245 DRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATD 304
          :: |: |: ::||:|:|:|||||:|  ||  || ||  | ||| ||   | |:  | |
Db    235 TNVSTVATEVETLSATVQEIASTADEVVDTSATAERLADDGSAAASDAADMMADVATAAD 294

Qy    305 GVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKA 364
          ||: || |  |  |:: |  ||  ||||||||||||||||||| |||||||:||||
Db    295 SVTSDVEALQNRIEDIDEVVDVITGIAEQTNMLALNASIEAARAGEEGEGFAVVAEEVKA 354

Qy    365 LAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVED 424
          |||:::  :  :| || ::| :| :||  |     || : | :||:|| : :||  |||
Db    355 LAEDAQSNAGHIESLVSEIQRDTADTVTDLVTTTDRIEDAVAQVEDAMASFEEIVTAVEA 414

Qy    425 AASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRET 484
          | |:::|| ||:||| |' ||:| |||   | | :|  |: ||   |: |   : :: |:
Db    415 TAEGIEQVSDATNEQAASAEEIAAMVDETADLADDITTAVADIVSQTEAQSAMLHDLDES 474

Qy    485 VGKL 488
          | :|
Db    475 VSEL 478


RESULT 4
T44973
transducer protein htpIII [similarity] - Halobacterium salinarum
N;Alternate names: methyl-accepting taxis protein htA; transducer protein htA; transd
C;Species: Halobacterium salinarum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Mar-2001
```

C;Accession: T44973
R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A;Title: Signal transduction in the archaeon Halobacterium salinarium is processed throu
A;Reference number: Z22804; MUID:96209786; PMID:8643458
A;Accession: T44973
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-482 <ZHA>
A;Cross-references: EMBL:U75435; NID:g1654418; PIDN:AAB17880.1; PID:g1654419
A;Experimental source: strain Flx15
A;Note: the source is designated as Halobacterium salinarium
C;Genetics:
A;Gene: htpIII; htp3; htA
C;Superfamily: Halobacterium salinarum transducer protein car
C;Keywords: methylated amino acid; signal transduction
F;219-472/Region: MCP signalling domain similarity

Query Match 25.5%; Score 609.5; DB 2; Length 482;
Best Local Similarity 37.2%; Pred. No. 6.3e-20;
Matches 181; Conservative 70; Mismatches 197; Indels 39; Gaps 12;

```
Qy      15 NGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADAL-VTDFYDHLE 73
           | :|  | |  :  | | :||        || :|||     |  |  ||: |||  : |
Db      19 NTLDVPAFA--VDADGAVVAW--------DDQIAAL-----LETAPEDAIGVTDIGERLN 63

Qy      74 SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVY 133
                 | : |  :   :   ||    | |  |  ||           :    |::
Db      64 D----DGSRALANKVADTPIDAHHEY---DGVGLADESYALLTGDYVYEDTTVAGNTDLW 116

Qy     134 LGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLL------TFDQQI--- 184
             |   |:||    :  ::| ||   :|    || | |  ::  |       ||  :
Db     117 FIATPVYHTGEFRGVI-EIVQDR-SSSARYQSELQALFGELVDTLDAYDAGRFDATVDIA 174

Qy     185 AMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQV 244
           | || :|    ::   :    :   |  | |   :  ||| || |:| :  :   :  |
Db     175 AEDTLLDDEYIQIGRNLTEFGDTLAAHITEVHNDVERLEAASQAVSESSAEIDELSTAQS 234

Qy     245 DRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATD 304
           :: |: |: ::||:|:|:|||||:|  ||  || ||  | ||| ||   | |:  | |
Db     235 TNVSTVATEVETLSATVQEIASTADEVVDTSATAERLADDGSAAASDAADMMADVATAAD 294

Qy     305 GVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKA 364
           ||: || |  |  |:: |  ||  ||||||||||||||||  || ||||||||:||||
Db     295 SVTSDVEALQNRIEDIDEVVDVITGIAEQTNMLALNASIEAAPPGEEGEGFAVVAEEVKA 354

Qy     365 LAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVE- 423
           |||:::  :  :| || ::| :| :||  |     || : | :||:|| : :||  |||
Db     355 LAEDAQSNAGHIESLVSEIQRDTADTVTDLVTTTDRIEDAVAQVEDAMASFEEIVTAVEA 414

Qy     424 --DAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEV 481
             : || |  || ||:||| | ||:| |||   | | :|  |: ||   |: |   : ::
Db     415 TAERASNM--VSDATNEQAASAEEIAAMVDETADLADDITTAVADIVSQTEAQSAMLHDL 472

Qy     482 RETVGKL 488
             |:| :|
Db     473 DESVSEL 479
```

RESULT 5
H84336
Htr3 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
C;Accession: H84336
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: H84336
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-633 <STO>
A;Cross-references: GB:AE004437; NID:g10581300; PIDN:AAG20060.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr3
C;Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 24.9%; Score 597; DB 2; Length 633;
Best Local Similarity 33.8%; Pred. No. 3e-19;
Matches 150; Conservative 76; Mismatches 158; Indels 60; Gaps 7;

```
Qy      42 IDDDTMAALAAEQPLFEATADALVTDFYDHLESYERT----QDLFANSTKTVEQLKETQA 97
           :|:|   |  |          :||:||  :  |  |:|    ::  | | :  |:|:   |
Db     213 LDEDLPGAFGASLSQMHTRLEALITDLDEAREDAEQTRKDAEEARAASERLNERLERRAA 272

Qy      98 EYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRG 157
                 ||  | ||  |              |  ||          | :|| ::
Db     273 ---------EYSDEMAAAAA----------------GDLTR-------RLDEDVDSEPM 299

Qy     158 EEAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEA 217
           :: | | ::::            | :  :    :  : : | :|       | :| |
Db     300 QDIAEAFNDMMG----------DVEATL-----AQVRSIADAVD-------AASTDVST 336

Qy     218 PLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED 277
             : : |  |:|    : |  | | ||:  |  |::|:||:|||:|::||||  |
Db     337 SAAEIRSASDQVSESVQDISADADQQRDRLGTVGDEVTSLSATVEEIAASADDVAETVNQ 396

Qy     278 AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNML 337
           |   :::|:   :||:| :  |:   |     |  | |    :  |||||  ||||||||
Db     397 AATESERGQELGEDAVAELERIEATADSAVERVTALEEAVDAIGDVTGVITDIAEQTNML 456

Qy     338 ALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEV 397
           ||||:|||||| ::|:|||||||||| ||:| :| :| :| ||: :||:  :||  : |:
Db     457 ALNANIEAARADKSGDGFAVVADEVKDLADEVKESATEIETLVDDVQADVADTVADMSEL 516

Qy     398 NQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRA 457
             |:  | | :| |:  | :| | || |   :| :| ||||| |||||  |:| | | :
Db     517 GDRVDAGSETIEAALAALDDIGDQVEAANGSVQSISDATDEQAASTEEVVTMIDEVTDLS 576

Qy     458 GEIAAALDDIADATDQQVRTVEEV 481
             |     :: | ::|  :| ||
Db     577 DRTATESQQVSAAAEEQAASVSEV 600
```

RESULT 6
T48840
transducer protein basT [validated] - Halobacterium salinarum
N;Alternate names: chemotaxis transducer protein basT; methyl-accepting taxis protein
C;Species: Halobacterium salinarum
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Sep-2000
C;Accession: T48840; T44981
R;Kokoeva, M.V.; Oesterhelt, D.
Mol. Microbiol. 35, 647-656, 2000
A;Title: BasT, a membrane-bound transducer protein for amino acid detection in Haloba
A;Reference number: Z24542; MUID:20138366; PMID:10672186
A;Accession: T48840
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-805 <KOK>
A;Cross-references: EMBL:AJ245950; PIDN:CAB82572.1
A;Experimental source: strain S9
R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A;Title: Signal transduction in the archaeon Halobacterium salinarium is processed th
A;Reference number: Z22804; MUID:96209786; PMID:8643458
A;Accession: T44981
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 'M',13-128,'R',130-133,'S',135-200,'MP',203-241,'R',243-332,'MR',335-401,
<ZHA>
A;Cross-references: EMBL:U75437; NID:g1654422; PIDN:AAB17882.1; PID:g1654423
A;Experimental source: strain Flx15 derivative of S9

```
A;Note: the source is designated as Halobacterium salinarium
C;Genetics:
A;Gene: basT; hpt10; htC
C;Function:
A;Description: involved in chemotactic signal transduction; mediates the response to fiv
C;Superfamily: Halobacterium salinarum transducer protein htrII
C;Keywords: methylated amino acid; signal transduction; transmembrane protein

  Query Match           24.9%;  Score 597;  DB 2;  Length 805;
  Best Local Similarity   33.8%;  Pred. No. 4e-19;
  Matches  150;  Conservative   76;  Mismatches  158;  Indels   60;  Gaps    7;

Qy      42 IDDDTMAALAAEQPLFEATADALVTDFYDHLESYERT----QDLFANSTKTVEQLKETQA 97
           :|:|   | |       :||:||  :  |  |:|    ::  | | :  |:|:   |
Db     385 LDEDLPGAFGASLSQMHTRLEALITDLDEAREDAEQTRKDAEEARAASERLNERLERRAA 444

Qy      98 EYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRG 157
                 ||  | ||  |                  |  ||        | :|| ::
Db     445 ---------EYSDEMAAAAA----------------GDLTR-------RLDEDVDSEPM 471

Qy     158 EEAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEA 217
           :: | | ::::           | :  :     :  : : | :|       | :| |
Db     472 QDIAEAFNDMMG----------DVEATL-----AQVRSIADAVD-------AASTDVST 508

Qy     218 PLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED 277
             : : :  |  ||:|    : |  | | ||:  |  |::|:||:|||:|::||||  |
Db     509 SAAEIRSASDQVSESVQDISADADQQRDRLGTVGDEVTSLSATVEEIAASADDVAETVNQ 568

Qy     278 AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNML 337
           |   :::|:   :||:| :  |:   |      |  | |    :  ||||| ||||||||
Db     569 AATESERGQELGEDAVAELERIEATADSAVERVTALEEAVDAIGDVTGVITDIAEQTNML 628

Qy     338 ALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEV 397
           ||||:|||||| ::|:|||||||||| ||:| :| :| :| ||: :||:  :||  : |:
Db     629 ALNANIEAARADKSGDGFAVVADEVKDLADEVKESATEIETLVDDVQADVADTVADMSEL 688

Qy     398 NQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRA 457
             |:  | | :| |:  | :| | || |   :| :| |||||| |||||  |:| | | :
Db     689 GDRVDAGSETIEAALAALDDIGDQVEAANGSVQSISDATDEQAASTEEVVTMIDEVTDLS 748

Qy     458 GEIAAALDDIADATDQQVRTVEEV 481
               |     :: | ::|  :| ||
Db     749 DRTATESQQVSAAAEEQAASVSEV 772


RESULT 7
B84238
Htr18 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: B84238
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: B84238
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-790 <STO>
A;Cross-references: GB:AE004437; NID:g10580382; PIDN:AAG19270.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr18
C;Superfamily: Halobacterium salinarum transducer protein htrII

  Query Match           24.2%;  Score 580;  DB 2;  Length 790;
  Best Local Similarity   31.4%;  Pred. No. 2.2e-18;
  Matches  160;  Conservative   84;  Mismatches  190;  Indels   76;  Gaps   11;

Qy       1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDD--DTMAALA---AEQP 55
             ||:  : :   |:   |:     | :|       | ||  | :   |:|  |   |||

Db     286 MSDRAEAIAAGDIDTAIEETTRIDEVG------DLRRSFRDIQEYLQTVAGQADALAEQD 339

Qy      56 LFEATA----------DALVTDFYD---HLESYERTQDLFANSTKTVEQLKETQAEYLLG 102
             |:| |          ::| |  :|    :   |  |:    | |  || :| :|| |
Db     340 -FDADALDKSVPGRLGESLETMHWDLETAIADLEDAQETAEQSRKEAEQSRE-EAEALA- 396

Qy     103 LGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAA 162
                     ||  ::   | : :   |                  |   :  |   |: |
Db     397 ----------AALESQAQDIRETVEHAAD-------------GDLTQRLETDTDHESMA 432

Qy     163 AVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVA---THVEAPL 219
           |:                      | ::|  : |   |   |  :  ||    |:
Db     433 AI----------------------ATALNSLLEELEGTIHRIQRFSKDVAESSDHITTSA 470

Qy     220 SSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAE 279
             ::  |  |:|    | |    |   : ||| |:: :||::||:||::|:|   | ||
Db     471 EEVKRASGQVSESVQEMSADARQQNGIVQDVSDEMTDLSATIEEIASSSDEVAAKSNDAV 530

Qy     280 ALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLAL 339
           :: | | : :||:  |  :||      | :| | :   ::  :  :|||||||||:||||
Db     531 SVGQSGRARSQDAIEEMNAVDEQAKRTIAEMEALDDEMTEIGEIVTLIDDIAEQTSMLAL 590

Qy     340 NASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQ 399
           |||||||||||||||||||||||:|:|::|: | :  :| |:  :|  | : |  : |:
Db     591 NASIEAARAGEAGEGFAVVADEIKSLSKETTEATQEIESLIADVQDSTTDAVTDMQEMGD 650

Qy     400 RIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGE 459
           |: ||   | : :|||:  | : :|:|  |:| ::|||||||| :||||  ||| |   : :
Db     651 RLSEGKSTVTDTVETIDTIVERIEEANGGVQTINTATDEQATTTEEVVTMVDEVGSISDD 710

Qy     460 IAAALDDIADATDQQVRTVEEVRETVGKLS 489
             |  :| | | ::|  :: ||   :  ||
Db     711 TTARAEDAAAAAEEQTASLTEVTNRIQDLS 740


RESULT 8
F84237
Htr4 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 24-Aug-2001
C;Accession: F84237
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: F84237
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-778 <STO>
A;Cross-references: GB:AE004437; NID:g10580376; PIDN:AAG19266.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr4
C;Superfamily: Halobacterium salinarum transducer protein htrII

  Query Match           24.1%;  Score 578;  DB 2;  Length 778;
  Best Local Similarity   33.7%;  Pred. No. 2.6e-18;
  Matches  145;  Conservative   75;  Mismatches  172;  Indels   38;  Gaps    6;

Qy      88 TVEQLKETQAEYLLGLGRGEYDTEYAAQRA-RIGKIHDVLGLGPDVYLGAYTRYYTGLLD 146
           ||  | : :|:    | ||||||:    |   :|:: :      |             :|
Db     320 TVRALNDLEAK-AAALERGEYDTDLDVARVDELGRLFEAFASLRDTVQARIRDANEQQVD 378

Qy     147 ----------ALADDVVADRGEEAAAAVDELVARFL---------PMLKLLTFDQQIAMD 187
                     | ||   |    |||    |  || |         |      | : :|
Db     379 AEAARSEAEAAQADAEAAQAEAEAAREESEAQARRLETTAEAFSETMRAYAAGDLTVRLD 438

Qy     188 TYIDSYAQ----------------RLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAE 231
             ::  |                 : | :    |:| | :|      :::| | :||::
Db     439 ADVEQAAMADIAAAFNEMAADMEATIADVVAFADEVATA-STDASDSAAAVEQTGRDVSD 497
```

```
Qy    232 RTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADD 291
Db    498 AVGRIRDRAADQRDQLEAVASETDEMSATIEEVAASADQVAETSQRAAALGDDGQAAAQD 557

Qy    292 ALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEA 351
Db    558 AVAQLEEIEDETQAAATAVDDLEAKMSEIETIVAAITDIAEQTNMLALNANIEAARADQD 617

Qy    352 GEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEA 411
Db    618 GDGFAVVADEVKDLADESKASAAEIEALVAEVRAQTETSVAAMDRIQERVSDGVETVSET 677

Qy    412 METLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADAT 471
Db    678 ERSLSEIAGRIAEADTGVQEISNAMDDQAASVSDVTTAVGDVAALGEETATEAESTADAA 737

Qy    472 DQQVRTVEEV 481
Db    738 AEQATTLSDV 747


RESULT 9
T48897
transducer protein VI (htp6, htD) [imported] - Halobacterium salinarum
C;Species: Halobacterium salinarum
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C;Accession: T48897
R;Rudolph, J.; Nordmann, B.; Storch, K.F.; Gruenberg, H.; Rodewald, K.; Oesterhelt, D.
FEMS Microbiol. Lett. 139, 161-168, 1996
A;Title: A family of halobacterial transducer proteins.
A;Reference number: Z24094; MUID:96275896; PMID:8674984
A;Accession: T48897
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-778 <RUD>
A;Cross-references: EMBL:X95590; NID:g1435133; PIDN:CAA64843.1; PID:g1435134
A;Experimental source: strain S6
C;Genetics:
A;Gene: htpVI
C;Superfamily: Halobacterium salinarum transducer protein htrII

  Query Match             23.9%;  Score 573;  DB 2;  Length 778;
  Best Local Similarity   33.0%;  Pred. No. 4.3e-18;
  Matches  144;  Conservative   74;  Mismatches  168;  Indels   50;  Gaps    6;

Qy     88 TVEQLKETQAEYLLGLGRGEYDTEYAAQRA-RIGKIHDVLGLGPDVYLGAYTRYYTGLLD 146
Db    320 TVRALNDLEAK-AAALERGEYDTDLDVARVDELGRLFEAFASLRDTVQARIRDANEQQVD 378

Qy    147 ----------ALADDVVADRGEEAAAAVDELVARFL---------PMLKLLTFDQQIAMD 187
Db    379 AEAARSEAEAAQADAEAAQAEAEAAREESEAQARRLETTAEAFSETMRAYAAGDLTVRLD 438

Qy    188 TYIDSYAQ----------------------RLHDEIDSRQELANAVATHVEAPLSSLEAT 225
Db    439 ADVEQAAMADIAAAFNEMAADMEATIADAGRFADEVAT-------ASTDASDSAAAVEQT 491

Qy    226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
Db    492 GRDVSDAVGRIRDRAADQRDQLEAVASETDEMSATIEEVAASADQVAETSQRAAALGDDG 551

Qy    286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
Db    552 QAAAQDAVAQLEEIEDETQAAATAVDDLEAKMSEIETIVAAITDIAEQTNMLALNANIEA 611

Qy    346 ARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGV 405
Db    612 ARADQDGDGFAVVADEVKDLADESKASAAEIEALVAEVRAQTETSVAAMDRIQERVSDGV 671

Qy    406 ERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALD 465
Db    672 ETVSETERSLSEIAGRIAEADTGVQEISNAMDDQAASVSDVTTAVGDVAALGEETATEAE 731

Qy    466 DIADATDQQVRTVEEV 481
Db    732 STADAAAEQASTLSDV 747


RESULT 10
T44849
transducer protein car [validated] - Halobacterium salinarum (strain S9)
N;Alternate names: arginine transducer protein Car; transducer protein htp11; transdu
C;Species: Halobacterium salinarum
A;Variety: strain S9
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000
C;Accession: T44849
R;Storch, K.F.; Rudolph, J.; Oesterhelt, D.
EMBO J. 18, 1146-1158, 1999
A;Title: Car: a cytoplasmic sensor responsible for arginine chemotaxis in the archaeo
A;Reference number: Z22859; MUID:99164082; PMID:10064582
A;Accession: T44849
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-452 <STO>
A;Cross-references: EMBL:AJ132321; NID:g4469245; PIDN:CAB38318.1; PID:g4469246
A;Experimental source: strain S9
A;Note: the source is designated as Halobacterium salinarum
C;Genetics:
A;Gene: car; htp11; htrXI
C;Function:
A;Description: involved in chemotactic signal transduction; arginine sensor; mediates
C;Superfamily: Halobacterium salinarum transducer protein car
C;Keywords: methylated amino acid; signal transduction

  Query Match             23.4%;  Score 560.5;  DB 2;  Length 452;
  Best Local Similarity   31.3%;  Pred. No. 8e-18;
  Matches  159;  Conservative   88;  Mismatches  170;  Indels   91;  Gaps   13;

Qy     10 TADVRNGIDGHALADRI----------GLDEAEIA------WRLSFTGIDDDTMAALAA 52
Db      5 SSDMGGEATGEHLADELCEAYLGDNEDDGGDELQRLSRERDFWKHMFN--------QLVA 56

Qy     53 EQP--LFEATADALVTD----FYDHLESY------ERTQDLFANSTKTVEQLKETQAEYL 100
Db     57 EYPEGILITAADGTVTHWNERFSDHMKMARSDALGEDASDVFS----TAEE-SETLPEAV 111

Qy    101 LGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEA 160
Db    112 VRTGDTVEEEE----------PHDVPTDSLCQYHGVPLRAPTG-------DVVGSFG--V 152

Qy    161 AAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLS 220
Db    153 VPDISEKVKN----------------------QRELHD-------LHETVSSNVGEHLS 182

Qy    221 SLEATSQDVAERTDTMRARTDDQVDRMADVSRRISSVSASVEEVASTADDVRRTSEDAEA 280
Db    183 ELSESIDEVGSFAEETEAFAGEEIERMEGVADEVSNQSATIEEIASSAEEVSQASQRAQD 242

Qy    281 LAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALN 340
Db    243 RATEGEQTAETAIDRMGAVQESAERVNDTIDGLTSQADEMSEIIDAINDIADQTNMLALN 302

Qy    341 ASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQR 400
Db    303 ASIEAARAGEKGEGFAVVADEVKSLAEESQERADEIEQMIVEMVETTDQTADRIGQTTTE 362

Qy    401 IGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEI 460
Db    363 IEEAITAVRETLDSLQEIRKAVDETATGVKEVAGATDDHAASTEQVAATTDEAVDKLTEL 422

Qy    461 AAALDDIADATDQQVRTVEEVRETVGKL 488
Db    423 EDRLDNLSQIASEQHDRVAEIEDMVDEL 450


RESULT 11
```

F84327
Htr5 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: F84327
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: F84327
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-810 <STO>
A;Cross-references: GB:AE004437; NID:g10581215; PIDN:AAG19986.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr5
C;Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 23.2%; Score 555; DB 2; Length 810;
Best Local Similarity 32.1%; Pred. No. 2.8e-17;
Matches 142; Conservative 85; Mismatches 174; Indels 42; Gaps 8;

```
Qy      84 NSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQR--------ARIGKIHDVLGLGPDVYLG 135
           |:   |: |   ||    :  | || : |: |        | || : | |    |
Db     346 NTAAAVQSLSAAAAE----IEAGNYDVDVASSRRDEIGQLFASIGSMRDALVTQIDEAEA 401

Qy     136 AYTRYYTGLLDALAD----DVVADRGEEAAAAVDELVA-------RFLPMLKLL------ 178
           |  :      || |:    :   :| |:| |  : | |       |:  ::
Db     402 AREQATEAQQDAEAERERAEDARERAEDAKADAEALAAELEAQAERYSDVMAACADGDLT 461

Qy     179 -------TFDQQIA--MDTYIDSYAQRLHDEIDSRQELANAVAT---HVEAPLSSLEATS 226
                 | :: :|    :: :  ||  |  ||  || |:|||     |   :  |  |
Db     462 RRMPADDTDNEAMAAIAASFNEMLAQWEHTIIDI-QEFADAVATASEEAEVGAADAERAS 520

Qy     227 QDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGE 286
             |:|    :   |:| : :  || |:: :||::|||::|| |   |     :|: ||
Db     521 GQVSESVQEIAGAADEQRNMLDTVSGEMTDLSAAIEEVAASADSVAEHSHQTAEIARDGE 580

Qy     287 AAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
             |:||:     : || |     || | :: |::  :  :| ||||||||||||:||||
Db     581 QTAEDAIERSLTVQEAIDATVQNVEALDDQMAEISEIVDLISDIAEQTNMLALNANIEAA 640

Qy     347 RAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVE 406
           || ::|:|||||||||| ||||::| :  :|  : ::|::|  || :     :  :  |::
Db     641 RADKSGDGFAVVADEVKDLAEETQESAGDIERRITEVQSQTTATVAEARAAEESMDAGID 700

Qy     407 RVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDD 466
            ||| ::    :|:|  ::  :|:||:|  ||:|| ||||    | : | | :   |
Db     701 AVEEVVDAFTAVSDHADETDTGVQEISDTTDDQAASTEEAVSMTEEVADLSDSTAGEAQS 760

Qy     467 IADATDQQVRTVEEVRETVGKLS 489
           :: | ::|  ::| |: ::|  ||
Db     761 VSAAAEEQAASMSEISDSVESLS 783
```

RESULT 12
T46810
halobacterial transducer protein IV [imported] - Halobacterium salinarum
C;Species: Halobacterium salinarum
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 02-Sep-2000
C;Accession: T46810
R;Rudolph, J.; Nordmann, B.; Storch, K.F.; Gruenberg, H.; Rodewald, K.; Oesterhelt, D.
FEMS Microbiol. Lett. 139, 161-168, 1996
A;Title: A family of halobacterial transducer proteins.
A;Reference number: Z24094; MUID:96275896; PMID:8674984
A;Accession: T46810
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-810 <RUD>
A;Cross-references: EMBL:X95589; NID:g1435130; PIDN:CAA64841.1; PID:g1435131
A;Experimental source: strain S9
C;Genetics:
A;Gene: htpIV
C;Superfamily: Halobacterium salinarum transducer protein htrII

Query Match 23.1%; Score 554; DB 2; Length 810;
Best Local Similarity 32.1%; Pred. No. 3.1e-17;
Matches 142; Conservative 85; Mismatches 174; Indels 42; Gaps 8;

```
Qy      84 NSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQR--------ARIGKIHDVLGLGPDVYLG 135
           |:   |: |   ||    :  | || : |: |        | || : | |    |
Db     346 NTAAAVQSLSAAAAE----IEAGNYDVDVASSRRDEIGQLFASIGSMRDALVTQIDEAEA 401

Qy     136 AYTRYYTGLLDALAD----DVVADRGEEAAAAVDELVA-------RFLPMLKLL------ 178
           |  :      || |:    :   :| |:| |  : | |       |:  ::
Db     402 AREQATEAQQDAEAERERAEDARERAEDAKADAEALAAELEAQAERYSDVMAACADGDLT 461

Qy     179 -------TFDQQIA--MDTYIDSYAQRLHDEIDSRQELANAVAT---HVEAPLSSLEATS 226
                 | :: :|    :: :  ||  |  ||  || |:|||     |   :  |  |
Db     462 RRMPADDTDNEAMAAIAASFNEMLAQWEHTIIDI-QEFADAVATASEEAEVGAADAERAS 520

Qy     227 QDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGE 286
             |:|    :   |:| : :  || |:: :||::|||::|| |   |     :|: ||
Db     521 GQVSESVQEIAGAADEQRNMLDTVSGEMTDLSAAIEEVAASADSVAEHSHQTAEIARDGE 580

Qy     287 AAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
             |:||:     : || |     || | :: |::  :  :| ||||||||||||:||||
Db     581 QTAEDAIERSLSVQEAIDATVQNVEALDDQMAEISEIVDLISDIAEQTNMLALNANIEAA 640

Qy     347 RAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVE 406
           || ::|:|||||||||| ||||::| :  :|  : ::|::|  || :     :  :  |::
Db     641 RADKSGDGFAVVADEVKDLAEETQESAGDIERRITEVQSQTTATVAEARAAEESMDAGID 700

Qy     407 RVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDD 466
            ||| ::    :|:|  ::  :|:||:|  ||:|| ||||    | : | | :   |
Db     701 AVEEVVDAFTAVSDHADETDTGVQEISDTTDDQAASTEEAVSMTEEVADLSDSTAGEAQS 760

Qy     467 IADATDQQVRTVEEVRETVGKLS 489
           :: | ::|  ::| |: ::|  ||
Db     761 VSAAAEEQAASMSEISDSVESLS 783
```

RESULT 13
T44597
transducer protein htpVI [similarity] - Halobacterium salinarum
N;Alternate names: methyl-accepting transducer protein htD; transducer protein htp6
C;Species: Halobacterium salinarum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000
C;Accession: T44597
R;Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A;Title: Signal transduction in the archaeon Halobacterium salinarium is processed th
A;Reference number: Z22804; MUID:96209786; PMID:8643458
A;Accession: T44597
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-777 <ZHA>
A;Cross-references: EMBL:U75438; NID:g1654424; PIDN:AAB17883.1; PID:g1654425
A;Experimental source: strain mFlx15
A;Note: the source is designated as Halobacterium salinarium
C;Genetics:
A;Gene: htpVI; htp6; htD
C;Superfamily: Halobacterium salinarum transducer protein htrII
C;Keywords: methylated amino acid; signal transduction; transmembrane protein
F;494-746/Region: MCP signalling domain similarity

Query Match 22.6%; Score 540.5; DB 2; Length 777;
Best Local Similarity 32.4%; Pred. No. 1.1e-16;
Matches 141; Conservative 74; Mismatches 171; Indels 49; Gaps 6;

```
Qy      88 TVEQLKETQAEYLLGLGRGEYDTEYAAQRA-RIGKIHDVLGLGPDVYLGAYTRYYTGLLD 146
           ||  | : :|:    | ||||||:    |   :|:: :      |    | :      :|
Db     320 TVRALNDLEAK-APALERGEYDTDLDVARVDELGRLFEAFASLRDAVQRADSDANEQQVD 378
```

```
Qy    147 ALA--------------------------------DDVVADRGEEAAAAVDELVARFLP 173
          | |                                ||     |::|  |  :|  |
Db    379 AEAARSEAEAAQAEPEAAQAEAEAAREESDAGAPPGDDRRGVLGDDACVAAGDLTVR--- 435

Qy    174 MLKLLTFDQQIA----MDTYIDSYAQRLHDEIDSRQELANAVA---THVEAPLSSLEATS 226
          |  | : |    |    :  |  :   |  |   |: ||   |      :::| |
Db    436 ----LDADVEQAAMADMRRAFNEMAADMEATIAERGAFADEVATASTDASDSAAAVEQTG 491

Qy    227 QDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGE 286
          :||:     :| |  || |::  |: |   :||::|||  |  |  ||: | ||   |:
Db    492 RDVSTAVGRIRDRAADQRDQLEAVASETDEMSATIEEVARVAGQVAETSQRAAALGDDGQ 551

Qy    287 AAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
          ||| ||:| : :|:: |      |: |  : :::|::   | ||||||||||||:||||
Db    552 AAAQDAVAQLEEIEDETQAAATAVDDLEAKMSEIETIVAAITDIAEQTNMLALNANIEAA 611

Qy    347 RAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVE 406
          || : |:|||||||||| ||:||:  :  :: || :::|:|| :|  :| : :|: :|||
Db    612 RADQDGDGFAVVADEVKDLADESKALAAEMQALVAEVRAQTETSVAAMDRIQERVSDGVE 671

Qy    407 RVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDD 466
           | |   :| :|   : :  :|:||:| | |:|| |  :|   |  |     | |   :
Db    672 TVSETERSLSDIAGRIAEPDTGVQEISNAMDDQAASVSDVTTAVGDVAALGEETATEAES 731

Qy    467 IADATDQQVRTVEEV 481
            ||  :|  |: :|
Db    732 TRDAAAEQATTLSDV 746


RESULT 14
S55299
sensory rhodopsin II transducer protein - Natronobacterium pharaonis
C;Species: Natronobacterium pharaonis
C;Date: 14-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Sep-2000
C;Accession: S55299
R;Seidel, R.; Scharf, B.; Gautel, M.; Kleine, K.; Oesterhelt, D.; Engelhard, M.
Proc. Natl. Acad. Sci. U.S.A. 92, 3036-3040, 1995
A;Title: The primary structure of sensory rhodopsin II: a member of an additional retina
A;Reference number: S55296; MUID:95224074; PMID:7708770
A;Accession: S55299
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-534 <SEI>
A;Cross-references: EMBL:Z35086; NID:g510868; PID:g510869
A;Experimental source: strain SP1
A;Note: it is uncertain whether Met-1 or Met-22 is the initiator
C;Genetics:
A;Gene: htrII
C;Superfamily: Halobacterium salinarum transducer protein htrI
C;Keywords: signal transduction; transmembrane protein
F;22-42/Domain: transmembrane #status predicted <TM1>
F;59-79/Domain: transmembrane #status predicted <TM2>

  Query Match             22.5%;  Score 537.5;  DB 2;  Length 534;
  Best Local Similarity   31.2%;  Pred. No. 9.8e-17;
  Matches  150;  Conservative   95;  Mismatches  177;  Indels   59;  Gaps   12;

Qy     16 GIDGHALADRIGLDEA----EIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFY-- 69
          ||:   :|  :| | |     :| : |  | | |     |       |   :  : | |
Db     72 GINLGLVAATLGGDTAASLSTLAAKASRMG-DGDLDVEL-------ETRREDEIGDLYAA 123

Qy     70 -DHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGL 128
           | :    ||    | : :   :  : :||       | :||  |:  | |:: |
Db    124 FDEMRQSVRTSLEDAKNAREDAEQAQKRAE--------EINTELQAEAERFGEVMDRCAD 175

Qy    129 GPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTFDQQIAMDT 188
          |           :|  |||  |:       ||  :::             :|::   ||
Db    176 GD----------FTQRLDAETDN-------EAMQSIEG-----------SFNEM--MDG 204

Qy    189 YIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMA 248
           |::   |:    |:  | | ||  : |   | :|| |:||      :      || : :
Db    205 -IEALVGRIERFADAVSEDAEAVRANAE---SVMEA-SEDVNRAVQNISDAAGDQTETVQ 259

Qy    249 DVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTA 308
          :: |:  |||: ||||::|||: :|:  |     : |    |: |:  | :::  |:   |
Db    260 QIALEMDDVSATTEEVAASADDIAKTARQAAETGEAGRETAETAITEMNEVESRTEQAVA 319

Qy    309 GVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEE 368
          :|:| |   ::  |: :| |||||||:|||||||||||    |||||||||||||||
Db    320 SMEELNEDVREIGEVSEMIADIAEQTNILALNASIEAARADGNSEGFAVVADEVKALAEE 379

Qy    369 SREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASG 428
          ::  :  :::|:  :|  |: ||| : | : :: |||| ||: :: |: | |:||    |
Db    380 TKAATEEIDDLIGTVQDRTQTTVDDIRETSDQVSEGVETVEDTVDALERIVDSVERTNDG 439

Qy    429 MQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKL 488
          :||:: :|| || : ::   ||: :   : : ||  :  |: |: |  :|:|| : :  |
Db    440 IQEINQSTDAQADAAQKATTMVEDMAATSEQTASDAETAAETTETQAESVKEVFDLIDGL 499

Qy    489 S 489
           |
Db    500 S 500


RESULT 15
A84252
Htr15 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84252
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: A84252
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-636 <STO>
A;Cross-references: GB:AE004437; NID:g10580513; PIDN:AAG19381.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr15

  Query Match             22.4%;  Score 535.5;  DB 2;  Length 636;
  Best Local Similarity   29.7%;  Pred. No. 1.5e-16;
  Matches  149;  Conservative   99;  Mismatches  191;  Indels   63;  Gaps   11;

Qy     17 IDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT----ADALVTD----- 67
          :|  || | ||:      : |  ||      : :|  : : | |    |||:  |
Db    165 VDDDALLDSIGMP----VFMLDATG------SVVAWNETIEELTGCSKADAMGMDHASEA 214

Qy     68 FYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLG 127
          ||      :  |      | :|  :  |        |:| | : ::|::     |:
Db    215 FYPDDRRVKTLAD------KVIESPRSAAA---------EFDIERSEEKAQLYADTSVM- 258

Qy    128 LGPDVYLGAYTRYYTGLLDALADD---------VVADRGEEA--AAAVDELVARFLPML- 175
                |   |:       : ||          : || |:   | ||:|||      :
Db    259 ----TDQGGTDRHIRFKASPIFDDDGELLAVAETIEDRTEDVRRADAVEELVDELSTTID 314

Qy    176 --------KLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQ 227
                  |  :|: :  ::  : |     |:  |  : |   |     :    ::|  :
Db    315 ALSSGQLSKRASFEHEGIINEQLVSVVSALNGMADQFERLVGQVDGQTQELADTIERATA 374

Qy    228 DVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEA 287
          |  :  ||:       | : ::  : |: : |||::|||:::| |   :| |:  |: |
Db    375 DATDIADTV----SSQNEMLSSAANEMENFSASMQEVAASSDQVASAAEQAQDAAESGLE 430

Qy    288 AADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAAR 347
          |:: |     :: : :| :   | :|  |  ::| |  || ::|:|||:|||||:|||||
Db    431 ASEGANQATNEVIDISDDLMESVSKLESRMDEIEDVVEVIAEVADQTNLLALNANIEAAR 490

Qy    348 AGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVER 407
          |||||  |||||||||| || |:|| : |:   :  :| :  |||  ::| :::|    :
```

```
Db    491 AGEAGSGFAVVADEVKELANETREHTERIAGSISDVQQQANETVLAVEESHEQIHRAGDE 550

Qy    408 VEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDI 467
          :::|:  |:|| :|::||:|: ||: | |||| : |:|   :: |  :| | ||| | |
Db    551 IDDALTALEEIATSVDEAATGITEVARANDEQASTVEDVIVTIEDVQQQAEEAAAASDRI 610

Qy    468 ADATDQQVRTVEEVRETVGKLS 489
            || :|   | |:: | | ||:
Db    611 VSATQEQSTAVSQLSERVDKLT 632


Search completed: January 2, 2003, 12:43:10
Job time : 23 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:39:17 ; Search time 35 Seconds
(without alignments)
2878.773 Million cell updates/sec

Title: US-09-455-978B-2
Perfect score: 2394
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVRTVEEVRETVGKLS 489

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	609.5	25.5	482	1	P71412	P71412 halobacteri
2	597	24.9	633	17	Q9HP10	Q9hp10 halobacteri
3	597	24.9	805	1	Q9P9J0	Q9p9j0 halobacteri
4	580	24.2	790	17	Q9HR88	Q9hr88 halobacteri
5	560.5	23.4	452	1	O93775	O93775 halobacteri
6	540.5	22.6	777	1	P71415	P71415 halobacteri
7	539	22.5	792	1	P71414	P71414 halobacteri
8	535.5	22.4	636	17	Q9HQX8	Q9hqx8 halobacteri
9	529	22.1	804	1	P71416	P71416 halobacteri
10	524	21.9	451	1	O06022	O06022 halobacteri
11	523.5	21.9	544	1	P71409	P71409 halobacteri
12	523.5	21.9	643	17	Q9HPQ5	Q9hpq5 halobacteri
13	521.5	21.8	628	17	Q9HRN6	Q9hrn6 halobacteri
14	509.5	21.3	789	17	Q9HRA1	Q9hra1 halobacteri
15	507.5	21.2	788	1	O93643	O93643 halobacteri
16	506	21.1	420	17	Q9HPW6	Q9hpw6 halobacteri
17	505	21.1	642	1	O59634	O59634 halobacteri
18	502.5	21.0	773	1	Q9UXS0	Q9uxs0 haloarcula
19	499.5	20.9	627	17	Q9HS86	Q9hs86 halobacteri
20	497.5	20.8	419	1	O93644	O93644 halobacteri
21	497	20.8	536	17	Q9HPA2	Q9hpa2 halobacteri
22	475.5	19.9	537	16	Q9AB06	Q9ab06 caulobacter
23	452.5	18.9	423	17	Q9HQT5	Q9hqt5 halobacteri
24	439.5	18.4	423	1	O93642	O93642 halobacteri
25	419.5	17.5	559	16	Q9K669	Q9k669 bacillus ha
26	415	17.3	555	16	Q9A5Y0	Q9a5y0 caulobacter
27	413	17.3	439	16	Q9KFH4	Q9kfh4 bacillus ha
28	409.5	17.1	632	16	Q9HW93	Q9hw93 pseudomonas
29	408	17.0	499	16	Q8UGL0	Q8ugl0 agrobacteri
30	405.5	16.9	579	16	Q9K632	Q9k632 bacillus ha
31	405.5	16.9	632	2	O32440	O32440 pseudomonas
32	401	16.8	530	16	Q9X0M7	Q9x0m7 thermotoga
33	401	16.8	629	16	Q9HW91	Q9hw91 pseudomonas
34	400.5	16.7	553	16	Q9KNF6	Q9knf6 vibrio chol
35	400	16.7	664	16	Q8RBV6	Q8rbv6 thermoanaer
36	399.5	16.7	535	16	Q9I0R3	Q9i0r3 pseudomonas
37	399	16.7	702	16	Q97D01	Q97d01 clostridium
38	397	16.6	566	16	Q9X1E2	Q9x1e2 thermotoga
39	396.5	16.6	541	16	Q9HUP8	Q9hup8 pseudomonas
40	396.5	16.6	834	17	O29217	O29217 archaeoglob
41	394	16.5	629	2	O32443	O32443 pseudomonas
42	393	16.4	667	16	Q9K617	Q9k617 bacillus ha
43	392.5	16.4	596	16	Q9KRJ7	Q9krj7 vibrio chol
44	391	16.3	667	2	Q9Z9T0	Q9z9t0 bacillus ha
45	390	16.3	623	16	Q9KQ43	Q9kq43 vibrio chol

ALIGNMENTS

RESULT 1
P71412
ID P71412 PRELIMINARY; PRT; 482 AA.
AC P71412;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Transducer HtA protein.
GN HTA.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FLX15;
RX MEDLINE=96209786; PubMed=8643458;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT "Signal transduction in the archaeon Halobacterium salinarium is
RT processed through three subfamilies of 13 soluble and membrane-bound
RT transducer proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
DR EMBL; U75435; AAB17880.1; -.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR SMART; SM00091; PAS; 1.
SQ SEQUENCE 482 AA; 50872 MW; ECBEF79B3374CEC7 CRC64;

Query Match 25.5%; Score 609.5; DB 1; Length 482;
Best Local Similarity 37.2%; Pred. No. 1.5e-18;
Matches 181; Conservative 70; Mismatches 197; Indels 39; Gaps 12;

Qy 15 NGIDGHALADRIGLDEAEIAWRLSFTGIDDDTNAALAAEQPLFEATADAL-VTDFYDHLE 73

```
Db      19 NTLDVPAFA--VDADGAVVAW--------DDQIAAL-----LETAPEDAIGVTDIGERLN 63

Qy      74 SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVY 133
                 | : |  :   :  ||    | |  |  ||          :    |::
Db      64 D----DGSRALANKVADTPIDAHHEY---DGVGLADESYALLTGDYVYEDTTVAGNTDLW 116

Qy     134 LGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLL------TFDQQI--- 184
             |   |:||   :  ::| ||   :|    || | |  ::  |        ||  :
Db     117 FIATPVYHTGEFRGVI-EIVQDR-SSSARYQSELQALFGELVDTLDAYDAGRFDATVDIA 174

Qy     185 AMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQV 244
           | || :|    :   :    :  | | |   :  ||| || |:| :  :   :  |
Db     175 AEDTLLDDEYIQIGRNLTEFGDTLAAHITEVHNDVERLEAASQAVSESSAEIDELSTAQS 234

Qy     245 DRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATD 304
           :: |: |: :||:|:|:|||||:|  ||  || ||  | ||| ||   | |:  | |
Db     235 TNVSTVATEVETLSATVQEIASTADEVVDTSATAERLADDGSAAASDAADMMADVATAAD 294

Qy     305 GVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKA 364
           ||: || |  ||  |:: |  ||  |||||||||||||||  || |||||||:||||
Db     295 SVTSDVEALQNRIEDIDEVVDVITGIAEQTNMLALNASIEAAPPGEEGEGFAVVAEEVKA 354

Qy     365 LAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVE- 423
           |||:::  :  :| || ::| :| :||  |     || : | :||:|| : :||  |||
Db     355 LAEDAQSNAGHIESLVSEIQRDTADTVTDLVTTTDRIEDAVAQVEDAMASFEEIVTAVEA 414

Qy     424 --DAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEV 481
             : || |  || ||:||| | ||:| |||   | | :|  |: ||   |: |   : ::
Db     415 TAERASNM--VSDATNEQAASAEEIAAMVDETADLADDITTAVADIVSQTEAQSAMLHDL 472

Qy     482 RETVGKL 488
             |:| :|
Db     473 DESVSEL 479
```

RESULT 2
Q9HP10
ID Q9HP10 PRELIMINARY; PRT; 633 AA.
AC Q9HP10;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Htr3 transducer.
GN HTR3 OR VNG1856G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005085; AAG20060.1; -.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
KW Complete proteome.
SQ SEQUENCE 633 AA; 66423 MW; C4D8022B66ECD0FF CRC64;

```
  Query Match             24.9%;  Score 597;  DB 17;  Length 633;
  Best Local Similarity   33.8%;  Pred. No. 7.1e-18;
  Matches  150;  Conservative   76;  Mismatches  158;  Indels   60;  Gaps    7;

Qy      42 IDDDTMAALAAEQPLFEATADALVTDFYDHLESYERT----QDLFANSTKTVEQLKETQA 97
           :|:|   | |          :||:||  :  |  |:|    ::  | | :  |:|:   |
Db     213 LDEDLPGAFGASLSQMHTRLEALITDLDEAREDAEQTRKDAEEARAASERLNERLERRAA 272

Qy      98 EYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRG 157
                 ||  | ||  |                  |  ||        | :|| ::
Db     273 ---------EYSDEMAAAAA----------------GDLTR-------RLDEDVDSEPM 299

Qy     158 EEAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEA 217
           :: | | ::::             | :  :     :  : : | :|       | :| |
Db     300 QDIAEAFNDMMG----------DVEATL-----AQVRSIADAVD-------AASTDVST 336

Qy     218 PLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED 277
           : : : |  |:|    : |  | | ||:  |  |::|:||:|||:|::||||  |
Db     337 SAAEIRSASDQVSESVQDISADADQQRDRLGTVGDEVTSLSATVEEIAAASADDVAETVNQ 396

Qy     278 AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNML 337
           |   :::|:   :||:| :  |:   |      |  | |    :  ||||| |||||||||
Db     397 AATESERGQELGEDAVAELERIEATADSAVERVTALEEAVDAIGDVTGVITDIAEQTNML 456

Qy     338 ALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEV 397
           ||||:|||||| ::|:|||||||||| ||:| :| :| :| ||: :||:  :||  : |:
Db     457 ALNANIEAARADKSGDGFAVVADEVKDLADEVKESATEIETLVDDVQADVADTVADMSEL 516

Qy     398 NQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRA 457
             |:  | | :| |:  | :| | || |   :| :| ||||||  |||||  |:| | | :
Db     517 GDRVDAGSETIEAALAALDDIGDQVEAANGSVQSISDATDEQAASTEEVVTMIDEVTDLS 576

Qy     458 GEIAAALDDIADATDQQVRTVEEV 481
             |     :: | ::|  :| ||
Db     577 DRTATESQQVSAAAEEQAASVSEV 600
```

RESULT 3
Q9P9J0
ID Q9P9J0 PRELIMINARY; PRT; 805 AA.
AC Q9P9J0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Chemotaxis transducer protein BasT.
GN BAST.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S9;
RX MEDLINE=20138366; PubMed=10672186;
RA Kokoeva M.V., Oesterhelt D.;
RT "BasT, a membrane-bound transducer protein for amino acid detection in
RT Halobacterium salinarum.";
RL Mol. Microbiol. 35:647-656(2000).
DR EMBL; AJ245950; CAB82572.1; -.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
SQ SEQUENCE 805 AA; 84831 MW; 089CA734D3F9DE3D CRC64;

```
  Query Match             24.9%;  Score 597;  DB 1;  Length 805;
  Best Local Similarity   33.8%;  Pred. No. 9.5e-18;
```

```
  Matches  150;  Conservative   76;  Mismatches  158;  Indels   60;  Gaps    7;

Qy      42 IDDDTMAALAAEQPLFEATADALVTDFYDHLESYERT----QDLFANSTKTVEQLKETQA 97
           :|:|   |  |          :||:||  :  |  |:|    ::  | | :  |:|:   |
Db     385 LDEDLPGAFGASLSQMHTRLEALITDLDEAREDAEQTRKDAEEARAASERLNERLERRAA 444

Qy      98 EYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRG 157
                 ||  | ||  |                  |  ||        | :|| ::
Db     445 ---------EYSDEMAAAAA----------------GDLTR-------RLDEDVDSEPM 471

Qy     158 EEAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEA 217
           :: | | ::::          | :  :     :  : : | :|       | :| |
Db     472 QDIAEAFNDMMG----------DVEATL-----AQVRSIADAVD-------AASTDVST 508

Qy     218 PLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED 277
             : : :  |:|    : |  | | ||:  |  |::|:||:|||:|::|||| |
Db     509 SAAEIRSASDQVSESVQDISADADQQRDRLGTVGDEVTSLSATVEEIAASADDVAETVNQ 568

Qy     278 AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNML 337
           |   :::|:   :||:| :  |:  |     |  | |    :  ||||| ||||||||
Db     569 AATESERGQELGEDAVAELERIEATADSAVERVTALEEAVDAIGDVTGVITDIAEQTNML 628

Qy     338 ALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEV 397
           ||||:|||||| ::|:|||||||||| ||:| :| :| :| ||: :||:  :||  : |:
Db     629 ALNANIEAARADKSGDGFAVVADEVKDLADEVKESATEIETLVDDVQADVADTVADMSEL 688

Qy     398 NQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRA 457
             |:  | | :| |:  | :| | || |   :| :| |||||| |||||  |:| | | :
Db     689 GDRVDAGSETIEAALAALDDIGDQVEAANGSVQSISDATDEQAASTEEVVTMIDEVTDLS 748

Qy     458 GEIAAALDDIADATDQQVRTVEEV 481
             |     :: | ::|  :| ||
Db     749 DRTATESQQVSAAAEEQAASVSEV 772


RESULT 4
Q9HR88
ID   Q9HR88      PRELIMINARY;      PRT;   790 AA.
AC   Q9HR88;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Htr18 transducer.
GN   HTR18 OR VNG0812G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005023; AAG19270.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Complete proteome.
SQ   SEQUENCE   790 AA;  84320 MW;  6227D876B5A35B85 CRC64;
```

```
  Query Match             24.2%;  Score 580;  DB 17;  Length 790;
  Best Local Similarity   31.4%;  Pred. No. 4.8e-17;
  Matches  160;  Conservative   84;  Mismatches  190;  Indels   76;  Gaps   11;

Qy       1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDD--DTMAALA---AEQP 55
           ||:  : :   |:   |:     | :|       | ||  | :   |:|  |   |||
Db     286 MSDRAEAIAAGDIDTAIEETTRIDEVG------DLRRSFRDIQEYLQTVAGQADALAEQD 339

Qy      56 LFEATA----------DALVTDFYD---HLESYERTQDLFANSTKTVEQLKETQAEYLLG 102
           |:| |          ::| |  :|    :    |  |:    | |  || :| :|| |
Db     340 -FDADALDKSVPGRLGESLETMHWDLETAIADLEDAQETAEQSRKEAEQSRE-EAEALA- 396

Qy     103 LGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAA 162
                 ||  ::   | : :   |                |   :  |   |: |
Db     397 ----------AALESQAQDIRETVEHAAD----------GDLTQRLETDTDHESMA 432

Qy     163 AVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVA---THVEAPL 219
           |:                      | ::|  : |   |   |  :  ||    |:
Db     433 AI----------------------ATALNSLLEELEGTIHRIQRFSKDVAESSDHITTSA 470

Qy     220 SSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAE 279
             ::  |  |:|    | |    |   : |||  :: :||::||:||::|:|   | ||
Db     471 EEVKRASGQVSESVQEMSADARQQNGIVQDVSDEMTDLSATIEEIASSSDEVAAKSNDAV 530

Qy     280 ALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLAL 339
           :: | | | : ||:  |  :||      | :|  :   ::  :  :|||||||:||||
Db     531 SVGQSGRARSQDAIEEMNAVDEQAKRTIAEMEALDDEMTEIGEIVTLIDDIAEQTSMLAL 590

Qy     340 NASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQ 399
           |||||||||||||||||||||:|:|::|: | :  :| |:  :|  | : |  : |:
Db     591 NASIEAARAGEAGEGFAVVADEIKSLSKETTEATQEIESLIADVQDSTTDAVTDMQEMGD 650

Qy     400 RIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGE 459
           |: ||   | : :||:  | : :|:|  |:| :|||||||| :|||  ||| |   : :
Db     651 RLSEGKSTVTDTVETIDTIVERIEEANGGVQTINTATDEQATTTEEVVTMVDEVGSISDD 710

Qy     460 IAAALDDIADATDQQVRTVEEVRETVGKLS 489
             |  :| | | ::|  :: ||   :  ||
Db     711 TTARAEDAAAAAEEQTASLTEVTNRIQDLS 740


RESULT 5
O93775
ID   O93775      PRELIMINARY;      PRT;   452 AA.
AC   O93775;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Car protein.
GN   CAR.
OS   Halobacterium salinarum.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S9;
RX   MEDLINE=99164082; PubMed=10064582;
RA   Storch K.F., Rudolph J., Oesterhelt D.;
RT   "Car: A cytoplasmic sensor responsible for arginine chemotaxis in the
RT   archaeon Halobacterium salinarum.";
RL   EMBO J. 18:1146-1158(1999).
DR   EMBL; AJ132321; CAB38318.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   InterPro; IPR000014; PAS_domain.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00283; MA; 1.
DR   SMART; SM00091; PAS; 1.
DR   TIGRFAMs; TIGR00229; sensory_box; 1.
```

```
SQ   SEQUENCE   452 AA;  49098 MW;  9244D900D694681A CRC64;

  Query Match           23.4%;  Score 560.5;  DB 1;  Length 452;
  Best Local Similarity   31.3%;  Pred. No. 1.6e-16;
  Matches  159;  Conservative   88;  Mismatches  170;  Indels   91;  Gaps   13;

Qy      10 TADVRNGIDGHALADRI----------GLDEAEIA------WRLSFTGIDDDTMAALAA 52
           ::|:     |  ||| :          | || :        |:  |          | |
Db       5 SSDMGGEATGEHLADELCEAYLGDNEDDGGDELQRLSRERDFWKHMFN--------QLVA 56

Qy      53 EQP--LFEATADALVTD----FYDHLESY------ERTQDLFANSTKTVEQLKETQAEYL 100
           | |  :    ||  ||     | ||::        |   |:|:    | |:  ||  | :
Db      57 EYPEGILITAADGTVTHWNERFSDHMKMARSDALGEDASDVFS----TAEE-SETLPEAV 111

Qy     101 LGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEA 160
           :  |    : |           |||       | |   |  ||       |||   |
Db     112 VRTGDTVEEEE----------PHDVPTDSLCQYHGVPLRAPTG-------DVVGSFG--V 152

Qy     161 AAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLS 220
           : | |                          : |||       |   |:::|   ||
Db     153 VPDISEKVKN----------------------QRELHD-------LHETVSSNVGEHLS 182

Qy     221 SLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEA 280
           |  :  :|   :   |   ::::||  |: |:|: ||::||:||:|::| : |: |:
Db     183 ELSESIDEVGSFAEETEAFAGEEIERMEGVADEVSNQSATIEEIASSAEEVSQASQRAQD 242

Qy     281 LAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALN 340
           | :||  |: |:  |  : |: : |   :: |  :| ::  :   |:|||:|||||||
Db     243 RATEGEQTAETAIDRMGAVQESAERVNDTIDGLTSQADEMSEIIDAINDIADQTNMLALN 302

Qy     341 ASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQR 400
           |||||||||| ||||||||||||:|||||:|::  :|::: :|   |::| |:: :
Db     303 ASIEAARAGEKGEGFAVVADEVKSLAEESQERADEIEQMIVEMVETTDQTADRIGQTTTE 362

Qy     401 IGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEI 460
           | | :  | | ::::||||  ||:: |:|::||: |||: | |||:||   |   |:  |:
Db     363 IEEAITAVRETLDSLQEIRKAVDETATGVKEVAGATDDHAASTEQVAATTDEAVDKLTEL 422

Qy     461 AAALDDIADATDQQVRTVEEVRETVGKL 488
              ||:::    :|   | |: : | :|
Db     423 EDRLDNLSQIASEQHDRVAEIEDMVDEL 450


RESULT 6
P71415
ID   P71415      PRELIMINARY;      PRT;   777 AA.
AC   P71415;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Transducer HtD protein.
GN   HTD.
OS   Halobacterium salinarum.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FLX15;
RX   MEDLINE=96209786; PubMed=8643458;
RA   Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT   "Signal transduction in the archaeon Halobacterium salinarium is
RT   processed through three subfamilies of 13 soluble and membrane-bound
RT   transducer proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
DR   EMBL; U75438; AAB17883.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF00015; MCPsignal; 1.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.

SQ   SEQUENCE   777 AA;  82097 MW;  0B20AA6A593AA816 CRC64;

  Query Match           22.6%;  Score 540.5;  DB 1;  Length 777;
  Best Local Similarity   32.4%;  Pred. No. 2.2e-15;
  Matches  141;  Conservative   74;  Mismatches  171;  Indels   49;  Gaps    6;

Qy      88 TVEQLKETQAEYLLGLGRGEYDTEYAAQRA-RIGKIHDVLGLGPDVYLGAYTRYYTGLLD 146
           ||  | : :|:    | ||||||:    |   :|:: :      |    | :      :|
Db     320 TVRALNDLEAK-APALERGEYDTDLDVARVDELGRLFEAFASLRDAVQRADSDANEQQVD 378

Qy     147 ALA--------------------------------DDVVADRGEEAAAAVDELVARFLP 173
           | |                                ||     |::|  |  :|  |
Db     379 AEAARSEAEAAQAEPEAAQAEAEAAREESDAGAPPGDDRRGVLGDDACVAAGDLTVR--- 435

Qy     174 MLKLLTFDQQIA----MDTYIDSYAQRLHDEIDSRQELANAVA---THVEAPLSSLEATS 226
           .     | | :|    |     :  | :   | |   |: ||   |       :::| |
Db     436 ----LDADVEQAAMADMRRAFNEMAADMEATIAERGAFADEVATASTDASDSAAAVEQTG 491

Qy     227 QDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGE 286
           :||:     :| |  || |::  |: |   :||:|||  |  |  ||: | ||   |:
Db     492 RDVSTAVGRIRDRAADQRDQLEAVASETDEMSATIEEVARVAGQVAETSQRAAALGDDGQ 551

Qy     287 AAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
           ||| ||:| : :|:: |       |: |  : :::|::   | ||||||||||||:||||
Db     552 AAAQDAVAQLEEIEDETQAAATAVDDLEAKMSEIETIVAAITDIAEQTNMLALNANIEAA 611

Qy     347 RAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVE 406
           || : |:|||||||||| ||:||:  :  :: || :::|:|| :|  :| : :|: :|||
Db     612 RADQDGDGFAVVADEVKDLADESKALAAEMQALVAEVRAQTETSVAAMDRIQERVSDGVE 671

Qy     407 RVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDD 466
           | |   :| :|   : :  :|:||:| | |:|| |  :|   |  |     | |   :
Db     672 TVSETERSLSDIAGRIAEPDTGVQEISNAMDDQAASVSDVTTAVGDVAALGEETATEAES 731

Qy     467 IADATDQQVRTVEEV 481
             ||  :|  |: :|
Db     732 TRDAAAEQATTLSDV 746


RESULT 7
P71414
ID   P71414      PRELIMINARY;      PRT;   792 AA.
AC   P71414;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transducer HtC protein.
GN   HTC.
OS   Halobacterium salinarum.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FLX15;
RX   MEDLINE=96209786; PubMed=8643458;
RA   Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT   "Signal transduction in the archaeon Halobacterium salinarium is
RT   processed through three subfamilies of 13 soluble and membrane-bound
RT   transducer proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
DR   EMBL; U75437; AAB17882.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR01608; BACINVASINC.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
SQ   SEQUENCE   792 AA;  83664 MW;  745D3693F3EEF9F9 CRC64;

  Query Match           22.5%;  Score 539;  DB 1;  Length 792;
```

```
  Best Local Similarity   32.2%;  Pred. No. 2.6e-15;
  Matches  143;  Conservative   76;  Mismatches  163;  Indels   62;  Gaps    7;

Qy      42 IDDDTMAALAAEQPLFEATADALVTDFYDHLESYERT----QDLFANSTKTVEQLKETQA 97
           :|:|   |  |          :||:||  :  |  |:|    ::  | | :  |:|:   |
Db     374 LDEDLPGAFGASLSQMH-RLEALITDLDEAREDAEQTRKDAEEARAASERLNERLERRAA 432

Qy      98 EYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRG 157
                 ||  | ||  |                  |  ||         | :|| :|
Db     433 ---------EYSDEMAAAAA-----------------GDLTR-------RLDEDVDSDGM 459

Qy     158 EEAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEA 217
           :: | | ::::          |  |  ::   :|: :  :||  : |
Db     460 QDIAEAFNDMMGDV-----EATLAQVRSIADAVDAASTDVHDAAEIR------------ 501

Qy     218 PLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED 277
             : |  |:|    : |  ::| ||:  |  |: ::||:||::|:        |
Db     502 ------SASDQVSESVQDISADAEEQRDRLGTVGDEVHALSATVEDIAARPTTSPSTVNQ 555

Qy     278 AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNML 337
           |   :::|:   :||:| :  |:   |      |  | |    :  |||| ||||||||
Db     556 AATESERGQELGEDAVAELERIEATADSAVERVTALEEAVDAIGHVTGVITDIAEQTNML 615

Qy     338 ALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEV 397
           ||||:|||||| ::|:||||||||||| ||:| :| :| :| ||: :||:  :||  : |:
Db     616 ALNANIEAARADKSGDGFAVVADEVKDLADEVKESATEIETLVDDVQADVADTVADMSEL 675

Qy     398 NQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRA 457
             |:  | | :| |:  : :| | || |   :| :| |||||| |||||  |:| | | :
Db     676 GDRVDAGSETIEAALAAVDDIGDQVEAANGSVQSISDATDEQAASTEEVVTMIDEVTDLS 735

Qy     458 GEIAAALDDIADATDQQVRTVEEV 481
             |     :: | ::|  :| ||
Db     736 DRTATESQQVSAAAEEQAASVSEV 759


RESULT 8
Q9HQX8
ID   Q9HQX8      PRELIMINARY;      PRT;   636 AA.
AC   Q9HQX8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Htr15 transducer.
GN   HTR15 OR VNG0958G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005032; AAG19381.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   InterPro; IPR000700; PAS-assoc_C.
DR   InterPro; IPR000014; PAS_domain.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00989; PAS; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00283; MA; 1.
DR   SMART; SM00091; PAS; 1.
DR   TIGRFAMs; TIGR00229; sensory_box; 1.
KW   Complete proteome.
SQ   SEQUENCE   636 AA;  67348 MW;  98D2E415295D4D14 CRC64;

  Query Match             22.4%;  Score 535.5;  DB 17;  Length 636;
  Best Local Similarity   29.7%;  Pred. No. 2.8e-15;
  Matches  149;  Conservative   99;  Mismatches  191;  Indels   63;  Gaps   11;

Qy      17 IDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT----ADALVTD----- 67
           :|  || | ||:      : |  ||      : :|  : : | |    |||:  |
Db     165 VDDDALLDSIGMP----VFMLDATG------SVVAWNETIEELTGCSKADAMGMDHASEA 214

Qy      68 FYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLG 127
           ||      :   |      | :|  :   |         |:| | : ::|::     |:
Db     215 FYPDDRRVKTLAD------KVIESPRSAAA---------EFDIERSEEKAQLYADTSVM- 258

Qy     128 LGPDVYLGAYTRYYTGLLDALADD---------VVADRGEEA--AAAVDELVARFLPML- 175
                 |   |:       : ||          : || |:   | ||:|||      :
Db     259 ----TDQGGTDRHIRFKASPIFDDDGELLAVAETIEDRTEDVRRADAVEELVDELSTTID 314

Qy     176 --------KLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQ 227
                   |  :|: :  ::  : |    |:  | : |   |    :    ::|  :
Db     315 ALSSGQLSKRASFEHEGIINEQLVSVVSALNGMADQFERLVGQVDGQTQELADTIERATA 374

Qy     228 DVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEA 287
           |  :  ||:       | : ::  : |: : ||| :|||:::| |   :| |:  |: |
Db     375 DATDIADTV----SSQNEMLSSAANEMENFSASMQEVAASSDQVASAAEQAQDAAESGLE 430

Qy     288 AADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAAR 347
           |:: |     :: : :| :   | :|  |  ::  |  || ::|:|||:|||||:|||||
Db     431 ASEGANQATNEVIDISDDLMESVSKLESRMDEIEDVVEVIAEVADQTNLLALNANIEAAR 490

Qy     348 AGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVER 407
           ||||| |||||||||| || |:|| : |:   :  :| :  |||  ::| :::|    :
Db     491 AGEAGSGFAVVADEVKELANETREHTERIAGSISDVQQQANETVLAVEESHEQIHRAGDE 550

Qy     408 VEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDI 467
           :::|:  |:||  :|::||:|: ||: | |||| : |:|   ::  |  :| | |||  | |
Db     551 IDDALTALEEIATSVDEAATGITEVARANDEQASTVEDVIVTIEDVQQQAEEAAAASDRI 610

Qy     468 ADATDQQVRTVEEVRETVGKLS 489
             || :|   | :: | | ||:
Db     611 VSATQEQSTAVSQLSERVDKLT 632


RESULT 9
P71416
ID   P71416      PRELIMINARY;      PRT;   804 AA.
AC   P71416;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Transducer HtF protein.
GN   HTF.
OS   Halobacterium salinarum.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FLX15;
RX   MEDLINE=96209786; PubMed=8643458;
RA   Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT   "Signal transduction in the archaeon Halobacterium salinarium is
RT   processed through three subfamilies of 13 soluble and membrane-bound
RT   transducer proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
DR   EMBL; U75439; AAB17884.1; -.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF00015; MCPsignal; 1.
DR   SMART; SM00304; HAMP; 2.
```

```
DR   SMART; SM00283; MA; 1.
SQ   SEQUENCE   804 AA;  84705 MW;  0D57284683BFD18B CRC64;

  Query Match           22.1%;  Score 529;  DB 1;  Length 804;
  Best Local Similarity   30.6%;  Pred. No. 6.9e-15;
  Matches  139;  Conservative   76;  Mismatches  159;  Indels   80;  Gaps    9;

Qy      89 VEQLKETQAEYLLGLGRGEYDTEYAAQR--------ARIGKIHDVLGLGPDVYLGAYTRY 140
           |: |    ||    :  | || : |  |         | || : | |    :    |  :
Db     351 VQSLSAAAAE----IEAGNYDVDVATSRRDEIGQLFASIGSMRDALVTQAE---AAREQA 403

Qy     141 YTGLLDALAD----DVVADRGEEAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQR 196
                 || ||    :   :| |:| |  : | |                  ::: |:|
Db     404 TEAQQDAEADARRAEDARERAEDAKADAEALAAE------------------LEAQAER 444

Qy     197 LHDEI-------------------------------DSR--------QELANAVAT---HV 215
             | |                               ||:        || |:||||
Db     445 YSDVIGGVADGDLTRRMPADDTDNEAMAAIAASFNDSQWEHTIIDIQEFADAVATASEEA 504

Qy     216 EAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTS 275
           |   :  | | |:|    :    |:| : :  || |:: :|||::||||| || |   |
Db     505 EVGAADAERASGQVSESVQEIAGAADEQRNMLDTVSGEMTDLSAAIEEVAS-ADSVAEHS 563

Qy     276 EDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTN 335
              :|: || | |:||:     : || |     || | :: |::  :  :| |||||||
Db     564 HQTAEIARDGEQTAEDAIERSLTVQEAIDATVQNVEALDDQMAEISEIVDLISDIAEQTN 623

Qy     336 MLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLD 395
           |||||:|||||| ::|:|||||||||||  |||::| :  :|  : ::|::|  || :
Db     624 MLALNANIEAARADKSGDGFAVVADEVKDFAEETQESAGDIERRITEVQSQTTATVAEAR 683

Qy     396 EVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDD 455
             : :  |:: ||| ::    ::|  ::  :|:||:|  ||:|| ||||   | : | |
Db     684 AAEESMDAGIDAVEEVVDAFTAVSDHSDETDTGVQEISDTTDDQAASTEEAVSMTEEVAD 743

Qy     456 RAGEIAAALDDIADATDQQVRTVEEVRETVGKLS 489
           :   |     ::   ::|  :: |: :||  ||
Db     744 LSDSTAGEAQSVSATAEEQAASMSEISDTVESLS 777


RESULT 10
O06022
ID   O06022      PRELIMINARY;      PRT;   451 AA.
AC   O06022;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Soluble transducer protein HtH.
OS   Halobacterium salinarum.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FLX15;
RX   MEDLINE=97284501; PubMed=9139915;
RA   Brooun A., Zhang W., Alam M.;
RT   "Primary structure and functional analysis of the soluble transducer
RT   protein HtrXI in the Archaeon Halobacterium salinarium.";
RL   J. Bacteriol. 179:2963-2968(1997).
DR   EMBL; U74668; AAC45264.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR000014; PAS_domain.
DR   Pfam; PF00015; MCPsignal; 1.
DR   SMART; SM00283; MA; 1.
DR   SMART; SM00091; PAS; 1.
DR   TIGRFAMs; TIGR00229; sensory_box; 1.
SQ   SEQUENCE   451 AA;  49122 MW;  6DA9E4FF8139DA87 CRC64;

  Query Match           21.9%;  Score 524;  DB 1;  Length 451;
  Best Local Similarity   30.9%;  Pred. No. 5.6e-15;
  Matches  158;  Conservative   87;  Mismatches  168;  Indels   98;  Gaps   16;

Qy      10 TADVRNGIDGHALADRI----------GLDEAEIA------WRLSFTGIDDDTMAALAA 52
           ::|:     |  ||| :           | || :        |:  |          | |
Db       5 SSDMGGEATGEHLADELCEAYLGDNEDDGGDELQRLSRERDFWKHMFN--------QLVA 56

Qy      53 EQP--LFEATADALVTD----FYDHLESY------ERTQDLFANSTKTVEQLKETQAEYL 100
           | |  :    ||  ||     | ||::        |   |:|:    | |:  ||  | :
Db      57 EYPEGILITAADGTVTHWNERFSDHMKMARSDALGEDASDVFS----TAEE-SETLPEAV 111

Qy     101 LGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEA 160
           :  |    : |           |||       | |   |  ||       |||   |
Db     112 VRTGDTVEEEE----------PHDVPTDSLCQYHGVPLRAPTG-------DVVGSFG--V 152

Qy     161 AAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLS 220
           : | |                              : |||       |   |: :|   ||
Db     153 VPDISEKVKN----------------------QRELHD-------LHETVSINVGEHLS 182

Qy     221 SLEATSQDV---AERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED 277
           |  :   |   || |:    :   :::||   : ::|: ||::||:||:|::| : |:
Db     183 ELSESIDKVGSFAEETEPFAGK---EIERMEGFADKVSNQSATIEEIASSAEEVSQASQR 239

Qy     278 AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNML 337
           |:  | :||  |: |:  |  : |: : |   :: |  :| ::  :   |:|||:|||||
Db     240 AQDRATEGEQTAETAIDRMGAVQESAERVNDTIDGLTSQADEMSEIIDAINDIADQTNML 299

Qy     338 ALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEV 397
           |||||||||||| ||||||||||||:|||||:|::  :|::: :|   |::| |:: :
Db     300 ALNASIEAARAGEKGEGFAVVADEVKSLAEESQERADEIEQMIVEMVETTDQTADRIGQT 359

Qy     398 NQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRA 457
           | | :  | | :::|||| :||:: |:|::||: | |  | |||:||   |   |:
Db     360 TTEIEEAITAVRETLDSLQEIRNAVDETATGVKEVAGARD-HAASTEQVAATTDEAVDKL 418

Qy     458 GEIAAALDDIADATDQQVRTVEEVRETVGKL 488
           |:   ||:::    :|   | |: : | :|
Db     419 TELEDRLDNLSQIASEQHDRVAEIEDMVDEL 449


RESULT 11
P71409
ID   P71409      PRELIMINARY;      PRT;   544 AA.
AC   P71409;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Transducer HtI protein.
GN   HTI.
OS   Halobacterium salinarum.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FLX15;
RX   MEDLINE=96209786; PubMed=8643458;
RA   Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT   "Signal transduction in the archaeon Halobacterium salinarium is
RT   processed through three subfamilies of 13 soluble and membrane-bound
RT   transducer proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
DR   EMBL; U53365; AAB17519.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   SMART; SM00304; HAMP; 1.
DR   SMART; SM00283; MA; 1.
SQ   SEQUENCE   544 AA;  56922 MW;  663D8D0E43AFCA71 CRC64;

  Query Match           21.9%;  Score 523.5;  DB 1;  Length 544;
  Best Local Similarity   29.4%;  Pred. No. 7.4e-15;
  Matches  141;  Conservative   93;  Mismatches  195;  Indels   51;  Gaps   10;
```

```
Qy      28 LDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDF----YDHLESYERTQDLFA 83
           |:|: :| :: |       ||   |   :      |:| |      :  :  :  ::  |
Db      64 LEKAGVATQI-FNLATGPLMATTVAVFAIGGTATLAIVEDMEALVEERAQPRQEAEEERA 122

Qy      84 NSTKTVEQLKETQAEYLLGLGRGEYDT-EYAAQRARIGKIHDVLGLGPDVYLGAYTRYYT 142
           : :  |: :: |||        |  | |  :|  | |   ::  |  |  | :
Db     123 EAERAREKAEQKQAE-------AERQTAEAQSQTGRTGTQREIEQLAAD--LESQATEVG 173

Qy     143 GLLDALAD-------DVVADRGE--EAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSY 193
             |:| :|       |   |  |  | |  |:::         | | :: |
Db     174 ATLEAASDGDLTARVDATTDNAEIAEVATVVNDM---------LTTMERTI--------- 215

Qy     194 AQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSRE 253
                |||          :          ::  || |:|    : | |||| :::  |: |
Db     216 -----DEIQGFSTNVTTASREATGGAKEIQVASQTVSESVQEIAAGTDDQREQLESVAEE 270

Qy     254 ISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQL 313
           : | ||:|||||:||  |  |: |   :|  |:  |:||:  :  : |       | |: |
Db     271 MDSYSATVEEVAATAQSVADTAADTTDVATAGKQTAEDAIDAIDAVQETMQTTVANVDAL 330

Qy     314 GERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEA----GEGFAVVADEVKALAEES 369
             :   ::: :  :| ||||||||||||||:|||||||      |:||||||||| || ||
Db     331 EDLTTEIDDIAELISDIAEQTNMLALNANIEAARAGSGGGTNGDGFAVVADEVKELATES 390

Qy     370 REQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGM 429
           :  :  : ||:|::|::|  ||:::    ||: :|   ||| ::   :|: :::   |:
Db     391 QRSAKDIAELIEEVQSQTATTVEEIRVAEQRVNDGAAAVEETVDAFGAVTENIQETTDGV 450

Qy     430 QEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKLS 489
           ||:| | ||||  :| |   || :   :   |   :::: |:::|  :: ||  ::  |:
Db     451 QEISQAMDEQAQRSERVVSSVDDIATISQATADRAENVSAASEEQTASITEVTGSLQSLA 510


RESULT 12
Q9HPQ5
ID   Q9HPQ5      PRELIMINARY;      PRT;   643 AA.
AC   Q9HPQ5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Htr8 transducer.
GN   HTR8 OR VNG1523G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005065; AAG19812.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00304; HAMP; 1.
DR   SMART; SM00283; MA; 1.
KW   Complete proteome.
SQ   SEQUENCE   643 AA;  67290 MW;  1D53F2E18BA02481 CRC64;

  Query Match           21.9%;  Score 523.5;  DB 17;  Length 643;
  Best Local Similarity  28.3%;  Pred. No. 9e-15;

  Matches  143;  Conservative   92;  Mismatches  200;  Indels   71;  Gaps    6;

Qy       8 LVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTD 67
           ||      :|: |   |:|:    | |      : ||     :  ||           ||: :
Db     156 LVYVVFTHGVFGMINAERVYNHTAAINNPWVWGGIHGAFVLLLAG----------ALMAN 205

Qy      68 FYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYA------------- 113
           :|    | | :|     : :  :|::: :|      : | : | |
Db     206 WYSTERSREASQKRLEEARQKAQQVEDLEAR------QAEIEAEKAEAKRLKADAEEARE 259

Qy     114 ---AQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALAD-DVVADRGEEAAAAVDELVA 169
               ||:  :  ::: |    : |  |  |  | |    | ||  |     ||: :|::
Db     260 AAEAQQREVAALNERLEATANTYGAAMARAADGDLSVRLDPDVENDAMAAIAASFNEML- 318

Qy     170 RFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHV-------EAPLSSL 222
                                          || ::      |||: |       :| :  :
Db     319 ----------------------------DETETTIREIQAVASDVAAASEDADAGVVEI 349

Qy     223 EATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALA 282
           |  |  |:|    : |  |:| :::  || |:: :|||::|||::|| |   | :  |:|
Db     350 EDASGQVSETVQEIAAGADEQREKLETVSGEMTDLSAAIEEVAASADSVAERSHETAAVA 409

Qy     283 QQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNAS 342
           ||  |: |:|    :  | :     || | :: |::  :  :| |:|||||||||||;
Db     410 GDGEQTAEQAIADSRTVQSAVESTVQNVEALDDQLAEISEIVDLISDVAEQTNMLALNAN 469

Qy     343 IEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIG 402
           |||||| ::|:|||||||||| ||||:|  :  :| ||  : |: : || :    :: :
Db     470 IEAARADKSGDGFAVVADEVKDLAEETRASAGDIEALVADIDAQMQATVTEARTADESVQ 529

Qy     403 EGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAA 462
           : :  |:  ::    : :  |:  :|:||:|| ||:|| ||||   |:  | | :   ||
Db     530 DAISAVDAVVDAFGTVAENAEETDTGVQEISTTTDDQAASTEEAVSMIAEVSDISTATAA 589

Qy     463 ALDDIADATDQQVRTVEEVRETVGKL 488
               : | :||     : |    |
Db     590 DAQQASTAAEQQTTAAATISENTAAL 615


RESULT 13
Q9HRN6
ID   Q9HRN6      PRELIMINARY;      PRT;   628 AA.
AC   Q9HRN6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Htr16 transducer.
GN   HTR16 OR VNG0614G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005010; AAG19122.1; -.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00304; HAMP; 2.
```

```
DR   SMART; SM00283; MA; 1.
KW   Complete proteome.
SQ   SEQUENCE   628 AA;  66833 MW;  98D1ECD7004ACC4B CRC64;

  Query Match             21.8%;  Score 521.5;  DB 17;  Length 628;
  Best Local Similarity   28.5%;  Pred. No. 1.1e-14;
  Matches  129;  Conservative   93;  Mismatches  167;  Indels   64;  Gaps    7;

Qy      42 IDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTKTVEQLK---ETQAE 98
           :||:    |       :   :  : |  |  |  ||:|     | :  : |    | |||
Db     217 LDDEVPGTLGESLTAMQTDLETFIEDLTDAREEAERSQRDAEQSREEAQALADSLERQAE 276

Qy      99 YLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGE 158
              : |:   |:|  |                        |   |  ||  :
Db     277 --------RFSTQ--MQQAADG---------------------DLTQRLDTDVDRESLA 304

Qy     159 EAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAV--ATH-V 215
           : | | ::::                            :|   :|  |  ||||   || :
Db     305 QIATAFNDMLV-------------------------QLEALVDRIQATANAVDDRTHAM 338

Qy     216 EAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTS 275
            |    :| :|  ||:  : : |  ::| :::   : |:: :|||:|||:||:::| : :
Db     339 SASTDEIEQSSMSVAKSIEQISAGAENQNEKLTAAAGEMTDLSATVEEIASSSNNVAQQA 398

Qy     276 EDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTN 335
           |||  |: | |:|||  ||:| |  |:          :: | |    :  :  :|||| ||:
Db     399 EDAAAMGQDGQAAATDAIAEMDAIESKATQTVEEMQSLQEEVGRISDIVTMIDDIASQTD 458

Qy     336 MLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLD 395
           |||:||||||  |  :|:|||||||||||:|||::   :  || |: :::| |    : :
Db     459 MLAINASIEAENASTSGDGFAVVADEVKSLAEDTATATDEVESLIAEVEASTSALAEDMY 518

Qy     396 EVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDD 455
           |:  |:|:| : |: :    |::| | |||| :|:| :: ||::|| | :|   ||: |
Db     519 EMRDRVGDGQDTITDTAALLEDIVDRVEDANAGIQSINDATEDQAASVQEATRMVEEVAS 578

Qy     456 RAGEIAAALDDIADATDQQVRTVEEVRETVGKL 488
            : : |    ::|: | ::|   ::|: ::   |
Db     579 VSDQTAGEASNVSAAAEEQTSAIQEISQSAQSL 611


RESULT 14
Q9HRA1
ID   Q9HRA1      PRELIMINARY;      PRT;   789 AA.
AC   Q9HRA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Htr6 transducer.
GN   HTR6 OR VNG0793G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005021; AAG19257.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Complete proteome.
SQ   SEQUENCE   789 AA;  83926 MW;  ABE57215065BF4B5 CRC64;

  Query Match             21.3%;  Score 509.5;  DB 17;  Length 789;
  Best Local Similarity   30.6%;  Pred. No. 4.5e-14;
  Matches  144;  Conservative   80;  Mismatches  165;  Indels   81;  Gaps    8;

Qy      21 ALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQD 80
           :| |||   |||                 |   |    :| |: | ||  |     || :
Db     361 SLQDRI--QEAE---------------AAVEEANAAKAEAEELRTDAEDAQAEAERAK- 401

Qy      81 LFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRY 140
             | :    |:|:|  |:
Db     402 --ATAEAASERLQERAAD---------------------------------------- 417

Qy     141 YTGLLDALADDVVADRGEEAA--AAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLH 198
           |: :: |:||  : :| :| |   |:  :   |  ||     |   |    :  :|  :
Db     418 YSEVMQAVADGDLTERLDEDADEEAMRAVATEFNAML-----DGLEATIAQVAGFADEVA 472

Qy     199 DEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVS 258
           ||          |||  |     :| ||| |:||   :       | | :   : |:  :|
Db     473 DE-------TLQVATGAE----EIETTSQTVSERIQEIADGAIQQHDDLERAAGEMDELS 521

Qy     259 ASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAA 318
           ||::|||:::  |  |: ||    : |  ||: |:  | :|:  :       :  | || :
Db     522 ASIQEVAASSATVAETAADAVERGEAGRDAAESAIDDMAEIESLSADAVDQILALQERMS 581

Qy     319 DVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEE 378
           |:  :   | ||||||||||||||:|||||| : |:|||||:||| ||||::: :  :|
Db     582 DIGDIIEFITDIAEQTNMLALNANIEAARADKDGDGFAVVANEVKDLAEETKQAAADIES 641

Qy     379 LVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDE 438
           :: :|||:|||  :   :: | :||  || |   :::  ||:||| |:||:| ||::
Db     642 EIQAVQAETDETVADIRATSEHIDDGVSTVERAAAAIEDTVDAIEDANHGIQEISDATED 701

Qy     439 QAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKL 488
           || :|: |   || | | :  :    : :: | ::|  :| |:  :   |
Db     702 QADATQSVVRRVDDVADISQHVTEDAEQVSAAAEEQSASVAEIARSADDL 751


RESULT 15
O93643
ID   O93643      PRELIMINARY;      PRT;   788 AA.
AC   O93643;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Transducer HtrVI.
GN   HTRVI.
OS   Halobacterium salinarum.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brooun A., Villablanca F., Freitas T., Hou S., Alam M.;
RT   "Structure, Function and Evolution of Transducer Family from the
RT   Archaeon Halobacterium salinarium.";
RL   J. Bacteriol. 0:0-0(1997).
DR   EMBL; AF036231; AAD02052.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   Pfam; PF00672; HAMP; 1.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
```

```
SQ   SEQUENCE   788 AA;  83868 MW;  C0D0839DDDBFC6CE CRC64;

  Query Match             21.2%;  Score 507.5;  DB 1;  Length 788;
  Best Local Similarity   30.9%;  Pred. No. 5.4e-14;
  Matches  143;  Conservative   79;  Mismatches  160;  Indels   81;  Gaps    8;

Qy      21 ALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQD 80
           :| |||   |||                |   |    :| |: | ||  |     || :
Db     361 SLQDRI--QEAE---------------AAVEEANAAKAEAEELRTDAEDAQAEAERAK- 401

Qy      81 LFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRY 140
             | :    |:|:|  |:
Db     402 --ATAEAASERLQERAAD---------------------------------------- 417

Qy     141 YTGLLDALADDVVADRGEEAA--AAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLH 198
           |: :: |:||  : :| :| |   |:  :   |  ||     |   |    :  :|  :
Db     418 YSEVMQAVADGDLTERLDEDADEEAMRAVATEFNAML-----DGLEATIAQVAGFADEVA 472

Qy     199 DEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVS 258
           ||          |||  |     :| ||| |:||   :      | | :   : |:  :|
Db     473 DE-------TLQVATGAE----EIETTSQTVSERIQEIADGAIQQHDDLERAAGEMDELS 521

Qy     259 ASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAA 318
           ||::|||:::  |  |: ||    : |  ||: |:  | :|:  :       :  | || :
Db     522 ASIQEVAASSATVAETAADAVERGEAGRDAAESAIDDMAEIESLSADAVDQILALQERMS 581

Qy     319 DVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEE 378
           |:  :   | ||||||||||||||:|||||| : |:|||||:||| ||||::: :  :|
Db     582 DIGDIIEFITDIAEQTNMLALNANIEAARADKDGDGFAVVANEVKDLAEETKQAAADIES 641

Qy     379 LVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDE 438
            :: :||||:|||  :   :: | :||  || |   :::  ||:|||  |:||:| ||::
Db     642 EIQAVQAETDETVADIRATSEHIDDGVSTVERAAAAIEDTVDAIEDANHGIQEISDATED 701

Qy     439 QAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEV 481
           || :|: |   || | | :  :    : :: | ::|  :| |:
Db     702 QADATQSVVRRVDDVADISQHVTEDAEQVSAAAEEQSASVAEI 744


Search completed: January  2, 2003, 12:42:42
Job time : 38 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:36:46 ; Search time 14 Seconds
(without alignments)
1448.709 Million cell updates/sec

Title: US-09-455-978B-2
Perfect score: 2394
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVRTVEEVRETVGKLS 489

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2394	100.0	489	1	HMAT_HALSA	P71413 halobacteri
2	2360	98.6	489	1	HMAT_HALN1	Q9hpr6 halobacteri
3	643	26.9	481	1	HTR3_HALN1	Q9hq00 halobacteri
4	643	26.9	481	1	HTR3_HALSA	Q48316 halobacteri
5	578	24.1	778	1	HTR6_HALN1	Q9hr92 halobacteri
6	573	23.9	778	1	HTR6_HALSA	Q48319 halobacteri
7	555	23.2	810	1	HTR4_HALN1	Q9hp84 halobacteri
8	554	23.1	810	1	HTR4_HALSA	Q48317 halobacteri
9	537.5	22.5	534	1	HTR2_NATPH	P42259 natronomona
10	521	21.8	545	1	HTR5_HALN1	Q48318 halobacteri
11	492	20.6	433	1	HTR2_HALVA	P42258 haloarcula
12	464	19.4	763	1	HTR2_HALN1	Q9hp81 halobacteri
13	463.5	19.4	535	1	HTR1_HALN1	P33741 halobacteri
14	461	19.3	764	1	HTR2_HALSA	P71410 halobacteri
15	449.5	18.8	535	1	HTR1_HALSA	P33955 halobacteri
16	412.5	17.2	432	1	HMAT_BACSU	O07621 bacillus su
17	388.5	16.2	662	1	MCPB_BACSU	P39215 bacillus su
18	386.5	16.1	682	1	PILJ_PSEAE	P42257 pseudomonas
19	379	15.8	662	1	TLPB_BACSU	P39217 bacillus su
20	359.5	15.0	547	1	MCPC_SALTY	Q02755 salmonella
21	359.5	15.0	1000	1	Y041_SYNY3	Q55445 synechocyst
22	354.5	14.8	662	1	TLPA_BACSU	P39216 bacillus su
23	351	14.7	548	1	HLYB_VIBCH	P15492 vibrio chol
24	350.5	14.6	573	1	TLPC_BACSU	P39209 bacillus su
25	349	14.6	553	1	MCP2_ECOLI	P07017 escherichia
26	348.5	14.6	551	1	MCP1_ECOLI	P02942 escherichia
27	342.5	14.3	657	1	MCPA_CAUCR	Q00986 caulobacter
28	335.5	14.0	553	1	MCP2_SALTY	P02941 salmonella
29	334	14.0	661	1	MCPA_BACSU	P39214 bacillus su
30	333.5	13.9	654	1	MCPC_BACSU	P54576 bacillus su
31	326.5	13.6	546	1	MCP3_ECOLI	P05704 escherichia
32	323.5	13.5	668	1	DCRA_DESVH	P35841 desulfovibr
33	318.5	13.3	557	1	MCPS_ENTAE	P21822 enterobacte
34	318	13.3	533	1	MCPE_RHIME	Q52877 rhizobium m
35	313.5	13.1	557	1	CPS_CLOTM	Q02929 clostridium
36	311.5	13.0	512	1	MCPD_ENTAE	P21823 enterobacte
37	309	12.9	756	1	Y4SI_RHISN	P55652 rhizobium s
38	308	12.9	506	1	AER_ECOLI	P50466 escherichia
39	304	12.7	533	1	MCP4_ECOLI	P07018 escherichia
40	299.5	12.5	845	1	Y4FA_RHISN	P55439 rhizobium s
41	286	11.9	620	1	TCPI_VIBCH	P29486 vibrio chol
42	271.5	11.3	417	1	FRZC_MYXXA	P43500 myxococcus
43	202.5	8.5	955	1	KINL_LEICH	P46865 leishmania
44	196.5	8.2	1972	1	MYHB_HUMAN	P35749 homo sapien
45	195.5	8.2	1972	1	MYHB_RABIT	P35748 oryctolagus

ALIGNMENTS

RESULT 1
HMAT_HALSA
ID HMAT_HALSA STANDARD; PRT; 489 AA.
AC P71413;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heme-based aerotactic transducer hemAT (Transducer HtB protein).
GN HEMAT OR HTB.
OS Halobacterium salinarium.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Flx15;
RX MEDLINE=96209786; PubMed=8643458;
RA Zhang W., Brooun A., McCandless J., Banda P., Alam M.;
RT "Signal transduction in the archaeon Halobacterium salinarium is processed through three subfamilies of 13 soluble and membrane-bound transducer proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4649-4654(1996).
RN [2]
RP FUNCTION.
RX MEDLINE=20140131; PubMed=10676961;
RA Hou S., Larsen R.W., Boudko D., Riley C.W., Karatan E., Zimmer M., Ordal G.W., Alam M.;
RT "Myoglobin-like aerotaxis transducers in Archaea and Bacteria.";
RL Nature 403:540-544(2000).
CC -!- FUNCTION: Heme-containing signal transducer responsible for aerotaxis, the migratory response toward or away from oxygen.
CC -!- SUBUNIT: Homotetramer (Probable).
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U75436; AAB17881.1; -.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00283; MA; 1.
KW Transducer; Heme.
SQ SEQUENCE 489 AA; 52816 MW; 8457263FCF616BFF CRC64;

Query Match 100.0%; Score 2394; DB 1; Length 489;
Best Local Similarity 100.0%; Pred. No. 7e-97;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

Qy     61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120

Qy    121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180

Qy    181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRART 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRART 240

Qy    241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300

Qy    301 EATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 EATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360

Qy    361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420

Qy    421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEE 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEE 480

Qy    481 VRETVGKLS 489
          |||||||||
Db    481 VRETVGKLS 489


RESULT 2
HMAT_HALN1
ID   HMAT_HALN1     STANDARD;      PRT;   489 AA.
AC   Q9HPR6;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Heme-based aerotactic transducer hemAT.
GN   HEMAT OR HTR10 OR VNG1505G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC   -!- FUNCTION: Heme-containing signal transducer responsible for
CC       aerotaxis, the migratory response toward or away from oxygen (By
CC       similarity).
CC   -!- SUBUNIT: Homotetramer (Probable).
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005064; AAG19801.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Heme; Complete proteome.
SQ   SEQUENCE   489 AA;  52852 MW;  8C0AEF179667791E CRC64;

  Query Match             98.6%;  Score 2360;  DB 1;  Length 489;
  Best Local Similarity   99.0%;  Pred. No. 2e-95;
  Matches  484;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

Qy     61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120

Qy    121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180

Qy    181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRART 240
          |||||||||||||||||||||||||||||: |||||||||||||||||||||||| |
Db    181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVASDVEAPLSSLEATSQDVAERTDTMRAAT 240

Qy    241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300

Qy    301 EATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 EATDGVTAGVEQLRERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360

Qy    361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420

Qy    421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEE 480
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDDVDDRAGEIAAALDDIADATDQQVRTVEE 480

Qy    481 VRETVGKLS 489
          |||||||||
Db    481 VRETVGKLS 489


RESULT 3
HTR3_HALN1
ID   HTR3_HALN1     STANDARD;      PRT;   481 AA.
AC   Q9HQ00;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Halobacterial transducer protein III.
GN   HTR9 OR VNG1395G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
```

```
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC       TRANSDUCTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005058; AAG19717.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   InterPro; IPR000014; PAS_domain.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00304; HAMP; 1.
DR   SMART; SM00283; MA; 1.
DR   SMART; SM00091; PAS; 1.
DR   PROSITE; PS50112; PAS; 1.
KW   Transducer; Complete proteome.
FT   DOMAIN       10     81       PAS.
SQ   SEQUENCE   481 AA;  50773 MW;  B333B55DBFE206D1 CRC64;

  Query Match             26.9%;  Score 643;  DB 1;  Length 481;
  Best Local Similarity   37.6%;  Pred. No. 1.8e-21;
  Matches  182;  Conservative   72;  Mismatches  196;  Indels   34;  Gaps   10;

Qy      15 NGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADAL-VTDFYDHLE 73
           | :|  | |  :  | | :||         || :|||     |  |  ||: |||  : |
Db      19 NTLDVPAFA--VDADGAVVAW--------DDQIAAL-----LETAPEDAIGVTDIGERLN 63

Qy      74 SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVY 133
                 | : |  :   :  ||    | |  |  ||           :    |::
Db      64 D----DGSRALANKVADTPIDAHHEY---DGVGLADESYALLTGDYVYEDTTVAGNTDLW 116

Qy     134 LGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLL------TFDQQI--- 184
             |   |:||    :  ::| ||   :|    || | |  ::  |       ||  :
Db     117 FIATPVYHTGEFRGVI-EIVQDR-SSSARYQSELQALFGELVDTLDAYDAGRFDATVDIA 174

Qy     185 AMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQV 244
           | || :|    ::   :    :   |  | |   :  ||| || |:| :  :   :  |
Db     175 AEDTLLDDEYIQIGRNLTEFGDTLAAHITEVHNDVERLEAASQAVSESSAEIDELSTAQS 234

Qy     245 DRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATD 304
             :: |: |: ::||:|:|:|||||:|  ||  || ||  | ||| ||    | |:  | |
Db     235 TNVSTVATEVETLSATVQEIASTADEVVDTSATAERLADDGSAAASDAADMMADVATAAD 294

Qy     305 GVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKA 364
             ||: || |  |  |:: |  ||  |||||||||||||||||||| ||||||||:||||
Db     295 SVTSDVEALQNRIEDIDEVVDVITGIAEQTNMLALNASIEAARAGEEGEGFAVVAEEVKA 354

Qy     365 LAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVED 424
           |||:::  :  :| || ::| :| :||  |     || : | :||:|| : :||  |||
Db     355 LAEDAQSNAGHIESLVSEIQRDTADTVTDLVTTTDRIEDAVAQVEDAMASFEEIVTAVEA 414

Qy     425 AASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRET 484
             | |:::|| ||:||| | ||:| |||   | | :|  |: ||   |: |   : :: |:
Db     415 TAEGIEQVSDATNEQAASAEEIAAMVDETADLADDITTAVADIVSQTEAQSAMLHDLDES 474

Qy     485 VGKL 488
             | :|
Db     475 VSEL 478


RESULT 4
HTR3_HALSA
ID   HTR3_HALSA     STANDARD;      PRT;   481 AA.
AC   Q48316;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Halobacterial transducer protein III.
GN   HTR9 OR HTPIII.
OS   Halobacterium salinarium.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S9;
RX   MEDLINE=96275896; PubMed=8674984;
RA   Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA   Oesterhelt D.;
RT   "A family of halobacterial transducer proteins.";
RL   FEMS Microbiol. Lett. 139:161-168(1996).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC       TRANSDUCTION.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X95588; CAA64840.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR000014; PAS_domain.
DR   Pfam; PF00015; MCPsignal; 1.
DR   SMART; SM00304; HAMP; 1.
DR   SMART; SM00283; MA; 1.
DR   SMART; SM00091; PAS; 1.
DR   PROSITE; PS50112; PAS; 1.
KW   Transducer.
FT   DOMAIN       10     81       PAS.
SQ   SEQUENCE   481 AA;  50728 MW;  B283AA8733F7844A CRC64;

  Query Match             26.9%;  Score 643;  DB 1;  Length 481;
  Best Local Similarity   37.6%;  Pred. No. 1.8e-21;
  Matches  182;  Conservative   72;  Mismatches  196;  Indels   34;  Gaps   10;

Qy      15 NGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADAL-VTDFYDHLE 73
           | :|  | |  :  | | :||         || :|||     |  |  ||: |||  : |
Db      19 NTLDVPAFA--VDADGAVVAW--------DDQIAAL-----LETAPEDAIGVTDIGERLN 63

Qy      74 SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVY 133
                 | : |  :   :  ||    | |  |  ||           :    |::
Db      64 D----DGSRALANKVADTPIDAHHEY---DGVGLADESYALLTGDYVYEDTTVAGNTDLW 116

Qy     134 LGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLL------TFDQQI--- 184
             |   |:||    :  ::| ||   :|    || | |  ::  |       ||  :
Db     117 FIATPVYHTGEFRGVI-EIVQDR-SSSARYQSELQALFGELVDTLDAYDAGRFDATVDIA 174

Qy     185 AMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQV 244
           | || :|    ::   :    :   |  | |   :  ||| || |:| :  :   :  |
```

```
Db    175 AEDTLLDDEYIQIGRNLTEFGDTLAAHITEVHNDVERLEAASQAVSESSAEIDELSTAQS 234

Qy    245 DRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATD 304
          :: |: |: :||:|:|:|||||:|  ||  || ||  | ||| ||    | |:  | |
Db    235 TNVSTVATEVETLSATVQEIASTADEVVDTSATAERLADDGSAAASDAADMMADVATAAD 294

Qy    305 GVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKA 364
           ||: || |  |  |:: |  ||  |||||||||||||||||||| |||||||:||||
Db    295 SVTSDVEALQNRIEDIDEVVDVITGIAEQTNMLALNASIEAARAGEVGEGFAVVAEEVKA 354

Qy    365 LAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVED 424
          |||:::  :  :| || ::| :| :||  |     || : | :||:|| : :||  |||
Db    355 LAEDAQSNAGHIESLVSEIQRDTADTVTDLVTTTDRIEDAVAQVEDAMASFEEIVTAVEA 414

Qy    425 AASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRET 484
          | |:::|| ||:||| | ||:| |||   | | :|  |: ||   |: |   : :: |:
Db    415 TAEGIEQVSDATNEQAASAEEIAAMVDETADLADDITTAVADIVSQTEAQSTILXDLDES 474

Qy    485 VGKL 488
          | :|
Db    475 VSEL 478


RESULT 5
HTR6_HALN1
ID   HTR6_HALN1     STANDARD;      PRT;   778 AA.
AC   Q9HR92;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Halobacterial transducer protein VI.
GN   HTR4 OR HTPVI OR VNG0806G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC       TRANSDUCTION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE005022; AAG19266.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Transmembrane; Complete proteome.
FT   DOMAIN        1     26       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      27    47       POTENTIAL.
FT   DOMAIN        48    296      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      297   317      POTENTIAL.
FT   DOMAIN        318   778      CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   778 AA;  82076 MW;  D752278727A4FA79 CRC64;

  Query Match             24.1%;  Score 578;  DB 1;  Length 778;
  Best Local Similarity   33.7%;  Pred. No. 1.9e-18;
  Matches  145;  Conservative   75;  Mismatches  172;  Indels   38;  Gaps    6;

Qy     88 TVEQLKETQAEYLLGLGRGEYDTEYAAQRA-RIGKIHDVLGLGPDVYLGAYTRYYTGLLD 146
          ||  | : :|:    | ||||||:    |   :|:: :      |             :|
Db    320 TVRALNDLEAK-AAALERGEYDTDLDVARVDELGRLFEAFASLRDTVQARIRDANEQQVD 378

Qy    147 ----------ALADDVVADRGEEAAAAVDELVARFL---------PMLKLLTFDQQIAMD 187
                     | ||   |    |||    |  || |          |      |  : :|
Db    379 AEAARSEAEAAQADAEAAQAEAEAAREESEAQARRLETTAEAFSETMRAYAAGDLTVRLD 438

Qy    188 TYIDSYAQ----------------RLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAE 231
          ::  |                  : | :   |:| | :|      :::| | :||::
Db    439 ADVEQAAMADIAAAFNEMAADMEATIADVVAFADEVATA-STDASDSAAAVEQTGRDVSD 497

Qy    232 RTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADD 291
            :| |  || |::  |: |   :||::||||::|| |  ||: | ||   |:||| |
Db    498 AVGRIRDRAADQRDQLEAVASETDEMSATIEEVAASADQVAETSQRAAALGDDGQAAAQD 557

Qy    292 ALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEA 351
          |:| : :|:: |      |: |  : :::|::   | ||||||||||||:|||||| :
Db    558 AVAQLEEIEDETQAAATAVDDLEAKMSEIETIVAAITDIAEQTNMLALNANIEAARADQD 617

Qy    352 GEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEA 411
          |:|||||||||| ||:||:  :  :| || :::|:|| :|  :| : :|: :||| | |
Db    618 GDGFAVVADEVKDLADESKASAAEIEALVAEVRAQTETSVAAMDRIQERVSDGVETVSET 677

Qy    412 METLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADAT 471
            :| ||   : :| :|:||:| | |:|| |  :|   |  |     | |   :  |||
Db    678 ERSLSEIAGRIAEADTGVQEISNAMDDQAASVSDVTTAVGDVAALGEETATEAESTADAA 737

Qy    472 DQQVRTVEEV 481
            :|  |: :|
Db    738 AEQATTLSDV 747


RESULT 6
HTR6_HALSA
ID   HTR6_HALSA     STANDARD;      PRT;   778 AA.
AC   Q48319;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Halobacterial transducer protein VI.
GN   HTR4 OR HTPVI.
OS   Halobacterium salinarium.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S9;
RX   MEDLINE=96275896; PubMed=8674984;
RA   Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA   Oesterhelt D.;
RT   "A family of halobacterial transducer proteins.";
RL   FEMS Microbiol. Lett. 139:161-168(1996).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC       TRANSDUCTION.
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X95590; CAA64843.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 1.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Transmembrane.
FT   DOMAIN        1     26       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     27     47       POTENTIAL.
FT   DOMAIN       48    296       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    297    317       POTENTIAL.
FT   DOMAIN      318    778       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   778 AA;  82077 MW;  134C7D7F0A3334CD CRC64;

  Query Match             23.9%;  Score 573;  DB 1;  Length 778;
  Best Local Similarity   33.0%;  Pred. No. 3.1e-18;
  Matches  144;  Conservative   74;  Mismatches  168;  Indels   50;  Gaps    6;

Qy      88 TVEQLKETQAEYLLGLGRGEYDTEYAAQRA-RIGKIHDVLGLGPDVYLGAYTRYYTGLLD 146
           ||  | : :|:    | |||||:    |   :|:: :      |                :|
Db     320 TVRALNDLEAK-AAALERGEYDTDLDVARVDELGRLFEAFASLRDTVQARIRDANEQQVD 378

Qy     147 ----------ALADDVVADRGEEAAAAVDELVARFL---------PMLKLLTFDQQIAMD 187
                    | ||   |    |||    |  || |           |      |  : :|
Db     379 AEAARSEAEAAQADAEAAQAEAEAAREESEAQARRLETTAEAFSETMRAYAAGDLTVRLD 438

Qy     188 TYIDSYAQ----------------------RLHDEIDSRQELANAVATHVEAPLSSLEAT 225
           ::  |                        |  ||: :      :|      :::| |
Db     439 ADVEQAAMADIAAAFNEMAADMEATIADAGRFADEVAT-------ASTDASDSAAAVEQT 491

Qy     226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
           :||::    :| |  || |::  |: |   :||::||||::|| |  ||: | ||   |
Db     492 GRDVSDAVGRIRDRAADQRDQLEAVASETDEMSATIEEVAASADQVAETSQRAAALGDDG 551

Qy     286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
           :||| ||:| : :|:: |      |: |  : :::|::   | ||||||||||||:|||
Db     552 QAAAQDAVAQLEEIEDETQAAATAVDDLEAKMSEIETIVAAITDIAEQTNMLALNANIEA 611

Qy     346 ARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGV 405
           ||| : |:||||||||||| ||:||:  :  :| || :::|:|| :|  :| : :|: :||
Db     612 ARADQDGDGFAVVADEVKDLADESKASAAEIEALVAEVRAQTETSVAAMDRIQERVSDGV 671

Qy     406 ERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALD 465
           | | |   :| ||   : :| :|:|:|:| | |:|| |  :|   |  |     | |   :
Db     672 ETVSETERSLSEIAGRIAEADTGVQEISNAMDDQAASVSDVTTAVGDVAALGEETATEAE 731

Qy     466 DIADATDQQVRTVEEV 481
           |||  :|  |: :|
Db     732 STADAAAEQASTLSDV 747


RESULT 7
HTR4_HALN1
ID   HTR4_HALN1     STANDARD;      PRT;   810 AA.
AC   Q9HP84;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Halobacterial transducer protein IV.
GN   HTR5 OR HTPIV OR VNG1760G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC       TRANSDUCTION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE005080; AAG19986.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 1.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Transmembrane; Complete proteome.
FT   DOMAIN        1     38       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     39     59       POTENTIAL.
FT   DOMAIN       60    323       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    324    344       POTENTIAL.
FT   DOMAIN      345    810       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   810 AA;  85219 MW;  4BF36E4B7D22BD80 CRC64;

  Query Match             23.2%;  Score 555;  DB 1;  Length 810;
  Best Local Similarity   32.1%;  Pred. No. 1.9e-17;
  Matches  142;  Conservative   85;  Mismatches  174;  Indels   42;  Gaps    8;

Qy      84 NSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQR--------ARIGKIHDVLGLGPDVYLG 135
           |:   |: |    ||    :  | || : |: |          | || : | |    |
Db     346 NTAAAVQSLSAAAAE----IEAGNYDVDVASSRRDEIGQLFASIGSMRDALVTQIDEAEA 401

Qy     136 AYTRYYTGLLDALAD----DVVADRGEEAAAAVDELVA-------RFLPMLKLL------ 178
           |  :      || |:    :   :| |:| |  : | |       |:  ::
Db     402 AREQATEAQQDAEAERERAEDARERAEDAKADAEALAAELEAQAERYSDVMAACADGDLT 461

Qy     179 -------TFDQQIA--MDTYIDSYAQRLHDEIDSRQELANAVAT---HVEAPLSSLEATS 226
                 | :: :|    :: :  ||  |  ||  || |:|||     |   :  |  |
Db     462 RRMPADDTDNEAMAAIAASFNEMLAQWEHTIIDI-QEFADAVATASEEAEVGAADAERAS 520

Qy     227 QDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGE 286
             |:|    :    |:| : :  || |:: :||::||||::|| |   |     :|: ||
Db     521 GQVSESVQEIAGAADEQRNMLDTVSGEMTDLSAAIEEVAASADSVAEHSHQTAEIARDGE 580

Qy     287 AAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
             |:||:     : || |     || | :: |::  :  :| |||||||||||||:||||
Db     581 QTAEDAIERSLTVQEAIDATVQNVEALDDQMARISEIVDLISDIAEQTNMLALNANIEAA 640

Qy     347 RAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVE 406
           || ::|:|||||||||| ||||::| :  :|  : ::|::|  || :     : :  |::
Db     641 RADKSGDGFAVVADEVKDLAEETQESAGDIERRITEVQSQTTATVAEARAAEESMDAGID 700

Qy     407 RVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDD 466
           |||  ::    ::|  ::  :|:||:|  ||:|| ||||   | : | | :   |
Db     701 AVEEVVDAFTAVSDHADETDTGVQEISDTTDDQAASTEEAVSMTEEVADLSDSTAGEAQS 760

Qy     467 IADATDQQVRTVEEVRETVGKLS 489
           :: | ::|  :: |: ::|  ||
Db     761 VSAAAEEQAASMSEISDSVESLS 783
```

```
RESULT 8
HTR4_HALSA
ID   HTR4_HALSA     STANDARD;      PRT;   810 AA.
AC   Q48317;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Halobacterial transducer protein IV.
GN   HTR5 OR HTPIV.
OS   Halobacterium salinarium.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S9;
RX   MEDLINE=96275896; PubMed=8674984;
RA   Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA   Oesterhelt D.;
RT   "A family of halobacterial transducer proteins.";
RL   FEMS Microbiol. Lett. 139:161-168(1996).
CC   -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC       TRANSDUCTION.
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X95589; CAA64841.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 1.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Transmembrane.
FT   DOMAIN        1     38       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     39     59       POTENTIAL.
FT   DOMAIN       60    323       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    324    344       POTENTIAL.
FT   DOMAIN      345    810       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   810 AA;  85207 MW;  37B0F6046A39D9BA CRC64;

  Query Match           23.1%;  Score 554;  DB 1;  Length 810;
  Best Local Similarity   32.1%;  Pred. No. 2.1e-17;
  Matches  142;  Conservative   85;  Mismatches  174;  Indels   42;  Gaps    8;

Qy      84 NSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQR--------ARIGKIHDVLGLGPDVYLG 135
           |:   |: |   ||    :  | || : |: |         | || : | |    |
Db     346 NTAAAVQSLSAAAAE----IEAGNYDVDVASSRRDEIGQLFASIGSMRDALVTQIDEAEA 401

Qy     136 AYTRYYTGLLDALAD----DVVADRGEEAAAAVDELVA-------RFLPMLKLL------ 178
           |  :      || |:    :   :| |:| |  : | |       |:  ::
Db     402 AREQATEAQQDAEAERERAEDARERAEDAKADAEALAAELEAQAERYSDVMAACADGDLT 461

Qy     179 -------TFDQQIA--MDTYIDSYAQRLHDEIDSRQELANAVAT---HVEAPLSSLEATS 226
                 | :: :|    :: :  ||  |  ||  || |:||||     |   :  |  |
Db     462 RRMPADDTDNEAMAAIAASFNEMLAQWEHTIIDI-QEFADAVATASEEAEVGAADAERAS 520

Qy     227 QDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGE 286
             |:|    :   |:| : :  || |:: :||::||||::|| |   |     :|: ||
Db     521 GQVSESVQEIAGAADEQRNMLDTVSGEMTDLSAAIEEVAASADSVAEHSHQTAEIARDGE 580

Qy     287 AAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
           |:||:     : || |     || | :: |::  :  :| ||||||||||||:||||
Db     581 QTAEDAIERSLSVQEAIDATVQNVEALDDQMAEISEIVDLISDIAEQTNMLALNANIEAA 640

Qy     347 RAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVE 406
           || ::|:|||||||||| ||||::| :  :|  : ::|::|  || :     : :  |::
Db     641 RADKSGDGFAVVADEVKDLAEETQESAGDIERRITEVQSQTTATVAEARAAEESMDAGID 700

Qy     407 RVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDD 466
           ||| ::    ::|  ::  :|:||:|  ||:|| |||   | : | | :   |
Db     701 AVEEVVDAFTAVSDHADETDTGVQEISDTTDDQASSTEEAVSMTEEVADLSDSTAGEAQS 760

Qy     467 IADATDQQVRTVEEVRETVGKLS 489
           :: | ::|  :: |: ::|  ||
Db     761 VSAAAEEQAASMSEISDSVESLS 783


RESULT 9
HTR2_NATPH
ID   HTR2_NATPH     STANDARD;      PRT;   534 AA.
AC   P42259;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE   protein II) (MPP-II).
GN   HTR2 OR HTRII.
OS   Natronomonas pharaonis (Natronobacterium pharaonis).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Natronomonas.
OX   NCBI_TaxID=2257;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SP-1 / 28;
RX   MEDLINE=95224074; PubMed=7708770;
RA   Seidel R., Scharf B., Gautel M., Kleine K., Oesterhelt D.,
RA   Engelhard M.;
RT   "The primary structure of sensory rhodopsin II: a member of an
RT   additional retinal protein subgroup is coexpressed with its
RT   transducer, the halobacterial transducer of rhodopsin II.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3036-3040(1995).
CC   -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC       SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC       LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35086; CAA84468.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 1.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Photoreceptor; Transmembrane; Methylation.
FT   DOMAIN        1     21       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     22     42       POTENTIAL.
FT   DOMAIN       43     58       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     59     79       POTENTIAL.
FT   DOMAIN       80    534       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   534 AA;  56622 MW;  CBE8012CC5F278E8 CRC64;

  Query Match           22.5%;  Score 537.5;  DB 1;  Length 534;
  Best Local Similarity   31.2%;  Pred. No. 7.1e-17;
  Matches  150;  Conservative   95;  Mismatches  177;  Indels   59;  Gaps   12;

Qy      16 GIDGHALADRIGLDEA----EIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFY-- 69
           ||:   :|  :| | |     :| : |  | | |    |       |   :  : | |
```

```
Db      72 GINLGLVAATLGGDTAASLSTLAAKASRMG-DGDLDVEL-------ETRREDEIGDLYAA 123

Qy      70 -DHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGL 128
Db     124 FDEMRQSVRTSLEDAKNAREDAEQAQKRAE--------EINTELQAEAERFGEVMDRCAD 175

Qy     129 GPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTFDQQIAMDT 188
Db     176 GD----------FTQRLDAETDN-------EAMQSIEG-----------SFNEM--MDG 204

Qy     189 YIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMA 248
Db     205 -IEALVGRIERFADAVSEDAEAVRANAE---SVMEA-SEDVNRAVQNISDAAGDQTETVQ 259

Qy     249 DVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTA 308
Db     260 QIALEMDDVSATTEEVAASADDIAKTARQAAETGEAGRETAETAITEMNEVESRTEQAVA 319

Qy     309 GVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEE 368
Db     320 SMEELNEDVREIGEVSEMIADIAEQTNILALNASIEAARADGNSEGFAVVADEVKALAEE 379

Qy     369 SREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASG 428
Db     380 TKAATEEIDDLIGTVQDRTQTTVDDIRETSDQVSEGVETVEDTVDALERIVDSVERTNDG 439

Qy     429 MQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKL 488
Db     440 IQEINQSTDAQADAAQKATTMVEDMAATSEQTASDAETAAETTETQAESVKEVFDLIDGL 499

Qy     489 S 489
Db     500 S 500
```

RESULT 10
HTR5_HALN1
ID HTR5_HALN1 STANDARD; PRT; 545 AA.
AC Q48318; Q9HP85;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Halobacterial transducer protein V.
GN HTR7 OR HTPV OR VNG1759G.
OS Halobacterium sp. (strain NRC-1), and
OS Halobacterium salinarium.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091, 2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRC-1;
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=H.salinarium; STRAIN=S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterhelt D.;
RT "A family of halobacterial transducer proteins.";
RL FEMS Microbiol. Lett. 139:161-168(1996).
CC -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION.
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AE005080; AAG19985.1; -.
DR EMBL; X95589; CAA64842.1; -.
DR HSSP; P02942; 1QU7.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
KW Transducer; Transmembrane; Complete proteome.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
SQ SEQUENCE 545 AA; 57070 MW; FDD870389C2F428B CRC64;

Query Match 21.8%; Score 521; DB 1; Length 545;
Best Local Similarity 29.6%; Pred. No. 3.7e-16;
Matches 142; Conservative 94; Mismatches 195; Indels 48; Gaps 10;

```
Qy      28 LDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDF----YDHLESYERTQDLFA 83
Db      64 LEKAGVATQI-FNLATGPLMATTVAVFAIGGTATLAIVEDMEALVEERAQRRQEAEEERA 122

Qy      84 NSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTG 143
Db     123 EAERAREKAEQKQAE----AERQTAEAESAKQDAR-ERSAEIEQLAAD--LESQATEVGA 175

Qy     144 LLDALAD-------DVVADRGE--EAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYA 194
Db     176 TLEAASDGDLTARVDATTDNAEIAEVATVVNDM---------LTTMERTI---------- 216

Qy     195 QRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREI 254
Db     217 ----DEIQGFSTNVTTASREATAGAKEIQDASQTVSESVQEIAAGTDDQREQLESVAEEM 272

Qy     255 SSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLG 314
Db     273 DSYSATVEEVAATAQSVADTAADTTDVATAGKQTAEDAIDAIDAVQETMQTTVANVDALE 332

Qy     315 ERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEA----GEGFAVVADEVKALAEESR 370
Db     333 DLTTEIDDIAELISDIAEQTNMLALNANIEAARAGSGGGSNGDGFAVVADEVKELATESQ 392

Qy     371 EQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQ 430
Db     393 RSAKDIAELIEEVQSQTATTVEEIRVAEQRVNDGAAAVEETVDAFGAVTENIQETTDGVQ 452

Qy     431 EVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKLS 489
Db     453 EISQAMDEQAQRSERVVSSVDDIATISQATADRAENVSAASEEQTASITEVTSSLQSLA 511
```

RESULT 11
HTR2_HALVA
ID HTR2_HALVA STANDARD; PRT; 433 AA.
AC P42258;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE protein II) (MPP-II) (Fragment).
GN HTRII.
OS Haloarcula vallismortis.
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

```
OC   Halobacteriaceae; Haloarcula.
OX   NCBI_TaxID=28442;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29715;
RX   MEDLINE=95224074; PubMed=7708770;
RA   Seidel R., Scharf B., Gautel M., Kleine K., Oesterhelt D.,
RA   Engelhard M.;
RT   "The primary structure of sensory rhodopsin II: a member of an
RT   additional retinal protein subgroup is coexpressed with its
RT   transducer, the halobacterial transducer of rhodopsin II.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3036-3040(1995).
CC   -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC       SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC       LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35308; CAA84549.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   SMART; SM00304; HAMP; 1.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Photoreceptor; Transmembrane; Methylation.
FT   NON_TER       1      1
SQ   SEQUENCE   433 AA;  45935 MW;  90507B8897D943C0 CRC64;

  Query Match           20.6%;  Score 492;  DB 1;  Length 433;
  Best Local Similarity  29.9%;  Pred. No. 5.2e-15;
  Matches  127;  Conservative   86;  Mismatches  172;  Indels   40;  Gaps    5;

Qy      69 YDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGL 128
           :|:: :  |||   | : |   :  : ||:      | : ::|     |  |
Db      11 FDNMRANLRTQISEAETAKQEAEAAKEQAQ----AAREDVESERNEMEALTG-------- 58

Qy     129 GPDVYLGAYTRYYTGLLDALADD-----VVADRGEEAAAAVDELVARFLPMLKLLTFDQQ 183
               :|    : |:  ||| |:      |  |   :| | | | :   |  |:    | :
Db      59 ----HLELKAQQYSDALDAAANGDLTARVKTDSMNDAMAEVGEDINTTLDALEDTIADMK 114

Qy     184 IAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQ 243
             | |     | ::|  |                ::  |: |::  : :   | :|
Db     115 AFATNVIQS-----SDRVNSNAE-------------RVDRASKQVSKSINEIFEGTTEQ 155

Qy     244 VDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEAT 303
            : :   : |: ::||: ::|||:|  |  ||: |  : : |  || :|:| |: |:  |
Db     156 NEGLESAAAEMQNLSATAQQVASSAQQVADTSQSAAKVGEDGREAAQEAIAEMSAIEAET 215

Qy     304 DGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVK 363
           |     : | :   ::  : ||| | ||||||||||||||| |   ||||||||||:|
Db     216 GETVEEINALDDELDEIGEIVGVITSIVEQTNMLALNASIEAAHADGDGEGFAVVADEIK 275

Qy     364 ALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVE 423
            ||||::| :  :|  :| :| :  :||: ::  : || |||  ||| :: |: | :  |
Db     276 GLAEETKEAAADIEGRIEAIQEQAGDTVETMESTSTRITEGVSTVEETVDALETIVEYTE 335

Qy     424 DAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRE 483
           :  :|:||:  ||:||| : ::|   :| :   : : |   | :| |   |  ::||| :
Db     336 EVDTGIQEIDRATEEQARTAQDVMGTIDDLTTISQQTATEADTVAGAAQDQSASIEEVSD 395

Qy     484 TVGKL 488
           :  :|
Db     396 SATEL 400
```

```
RESULT 12
HTR2_HALN1
ID   HTR2_HALN1     STANDARD;      PRT;   763 AA.
AC   Q9HP81;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE   protein II) (MPP-II).
GN   HTR2 OR VNG1765G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC   -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY
CC       RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT
CC       CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005080; AAG19989.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 2.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Photoreceptor; Transmembrane; Methylation;
KW   Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     15       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     16     36       POTENTIAL.
FT   DOMAIN       37    277       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    278    297       POTENTIAL.
FT   DOMAIN      298    763       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   763 AA;  78911 MW;  CF7A8FF04DFF309A CRC64;

  Query Match           19.4%;  Score 464;  DB 1;  Length 763;
  Best Local Similarity  28.9%;  Pred. No. 1.5e-13;
  Matches  136;  Conservative   74;  Mismatches  219;  Indels   42;  Gaps    9;

Qy      37 LSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQ 96
           ||  |:   :    |  |  |  || :    |     |  | |    ::   ::::
Db     293 LSLIGVTVGSTTVTALRQ--FSRRADEMAAGDLDTDIDTSR-NDEFGTLAESFRSMRDSL 349

Qy      97 AEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADR 156
           :| |        | | |  ||   :  |      |              :|  :|  | |
Db     350 SESLT-------DAERATARAEDAR-EDAEQQRADA------------EAAREDAEAAR 388

Qy     157 --GEEAAAAVDELVARFLPMLKLLT---FDQQIAMDTYIDSYAQRLH------DEIDSRQ 205
             :| | |::   | :   |  :      :::      |: |:  |      |:|::
```

```
Db    389 KDAQETARALESAAADYEEALTAVADGDLTRRVDASRDHDAMARIGHALNDMLDDIETSV 448

Qy    206 ELANAVATHVEAPLSSLEA-------TSQDVAERTDTMRARTDDQVDRMADVSREISSVS 258
          | | : ||      :||         ||:   | :     :| ||: :|: |:  :|
Db    449 AAATAFSDHVSDAAQRVEADAGDAIDAGTDVSTAVDEISDGATEQTDRLHEVAGEVDDLS 508

Qy    259 ASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAA 318
          || |||| |   :  |:  | :    |  | :||: || |: :  :     :: |    |
Db    509 ASAEEVAETVASLADTAGQAASAVDDGRQATEDAVETMDDVADDAEAAADAMDALDSEMA 568

Qy    319 DVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEE 378
          |:  :  || |||:|||||||||||||||| |  |:|||||||||| ||||||: :  :|
Db    569 DIGEIVDVIADIADQTNMLALNASIEAARTGADGDGFAVVADEVKTLAEESRDAAEDIES 628

Qy    379 LVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDE 438
           :  :| :  :  |::   :  : :|   | :|   | ::   | |  :   |:  |||
Db    629 RLLALQGQVSDVADEMRATSDTVSDGRATVGDAATALDDVVSFVADTDTAAGEIRAATDR 688

Qy    439 QAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKLS 489
          || :   ||  || |   : | ||    :||:   |  |:  | :    |:
Db    689 QAHAASRVASAVDEVAGISQETAAQATAVADSAATQTDTLSSVDDAAADLA 739


RESULT 13
HTR1_HALN1
ID   HTR1_HALN1     STANDARD;      PRT;   535 AA.
AC   P33741; Q9HPF6;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sensory rhodopsin I transducer (HTR-I) (Methyl-accepting phototaxis
DE   protein I) (MPP-I).
GN   HTR1 OR HTRI OR VNG1659G.
OS   Halobacterium sp. (strain NRC-1), and
OS   Halobacterium halobium.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091, 2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRC-1;
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10; 350-377 AND 457-476.
RC   SPECIES=H.halobium; STRAIN=FLX5R;
RX   MEDLINE=93101637; PubMed=1465418;
RA   Yao V.J., Spudich J.L.;
RT   "Primary structure of an archaebacterial transducer, a
RT   methyl-accepting protein associated with sensory rhodopsin I.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:11915-11919(1992).
CC   -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC       SENSORY RHODOPSIN I (SR-I) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC       LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005075; AAG19913.1; -.
DR   EMBL; L05603; AAA72315.1; -.
DR   PIR; A47190; A47190.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 1.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Photoreceptor; Transmembrane; Methylation;
KW   Complete proteome.
FT   INIT_MET      0      0
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     28       POTENTIAL.
FT   DOMAIN       29     38       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     39     54       POTENTIAL.
FT   DOMAIN       55    535       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     265    265       METHYLATION.
FT   MOD_RES     272    272       METHYLATION.
FT   MOD_RES     279    279       METHYLATION.
FT   MOD_RES     463    463       METHYLATION.
FT   MOD_RES     472    472       METHYLATION.
SQ   SEQUENCE   535 AA;  56544 MW;  B9945E4F66A9D091 CRC64;

  Query Match             19.4%;  Score 463.5;  DB 1;  Length 535;
  Best Local Similarity   27.5%;  Pred. No. 1.1e-13;
  Matches  147;  Conservative   87;  Mismatches  205;  Indels   95;  Gaps   15;

Qy      2 SNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATA 61
          :||  : : |    || |  |     :: ||   :       : :|    || :
Db     31 TNDVPSTIVA----GIAG--LLTLGSINAAETVASIKEIAAQTERVANGNLEQEVTSTRT 84

Qy     62 DALVTDFYDHLESYERTQDLFANSTKTVEQ---LKETQAEYLLGLGRGEYDTEYAAQRA 117
          |    :|    :| |: :          :|:    |:|||||          : | | | |
Db     85 D----EFGSLADSIEQMRQSLRGRLNEMERTRADLEETQAE----AETAREEAEQAKQEA 136

Qy    118 RIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGE--EAAAAVDELVARFLPML 175
          : :                  | |   |     ||  ||||  |
Db    137 QAAERE-----------------ARELAATYQDTAKRYGETMEAAATGD---------- 168

Qy    176 KLLTFDQQIAMDT------YIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDV 229
                |:: :||       : :   :: |:: :       ||  :||    :  || |:
Db    169 ----LTQRVDVDTDHEAMETVGTAFNQMMDDLQATVRTVTTVADEIEAKTERMSETSADI 224

Qy    230 -AERTDTMRA------RTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALA 282
           |   ||: |      : :||    :   : ::  |||| ||:|:| ||:   |||    :
Db    225 EASAGDTVEAVSKIESQANDQRTELDSAADDVQQVSASAEEIAATIDDLASRSEDVATAS 284

Qy    283 QQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNAS 342
              ::  ||  |: |:   |     |||| :: |:: :  :  || || ||||||||||
Db    285 DAARDSSKSALDEMSSIETEVDDAVGQVEQLRDQVAEITDIVDVITDIGEQTNMLALNAS 344

Qy    343 IEAARAG--EAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQR 400
          |||||||    |:||:|||||||| ||||:::::  :  :||:: |:||:    : :   |
Db    345 IEAARAGGNADGDGFSVVADEVKDLAEETQDRANEIAAVVEKVTAQTEDVTASIQQTRTR 404

Qy    401 IGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQ-------AVSTEEVAEMVD-- 451
          :  | | ||  :  :: | |:: : :: : |:   | ||          | | | || : |
Db    405 VESGSETVESTLRDIRTIADSIAEVSNSIDEIQRTTSEQAETVQSTATSVERVAGLSDDT 464

Qy    452 -----------GVDDRAGEIAAALDDIADATDQQVRT-------VEEVRETVG 486
                      |  : | ||||:|:   :   :|:::         |  || |
Db    465 TALASDAESAVIGQRESAEEIAASLEQFQNTAVEQLQSRVASFTVATEDSETAG 518


RESULT 14
HTR2_HALSA
ID   HTR2_HALSA     STANDARD;      PRT;   764 AA.
AC   P71410;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE   Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis
DE   protein II) (MPP-II).
GN   HTR2 OR HTRII.
OS   Halobacterium salinarium.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Flx15;
RX   MEDLINE=96323203; PubMed=8710852;
RA   Zhang W., Brooun A., Mueller M.M., Alam M.;
RT   "The primary structures of the Archaeon Halobacterium salinarium blue
RT   light receptor sensory rhodopsin II and its transducer, a methyl-
RT   accepting protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).
CC   -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC       SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC       LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U62676; AAC44369.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 2.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Photoreceptor; Transmembrane; Methylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     15       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     16     36       POTENTIAL.
FT   DOMAIN       37    277       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    278    298       POTENTIAL.
FT   DOMAIN      299    764       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   764 AA;  79187 MW;  1E0D7B4E460FC588 CRC64;

  Query Match           19.3%;  Score 461;  DB 1;  Length 764;
  Best Local Similarity  28.7%;  Pred. No. 2e-13;
  Matches  135;  Conservative   75;  Mismatches  219;  Indels   42;  Gaps    9;

Qy      37 LSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQ 96
           ||  |:   :   |  |  |  || :     |      |  | |    ::   ::::
Db     294 LSLIGVTVGSTTVTALRQ--FSRRADEMAAGDLDTDIDTSR-NDEFGTLAESFRSMRDSL 350

Qy      97 AEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADR 156
           :| |       | | |  ||   :  |      |                :|  :|  | |
Db     351 SESLT-------DAERATARAEDAR-EDAEQQRADA------------EAAREDAEAAR 389

Qy     157 --GEEAAAAVDELVARFLPMLKLLT---FDQQIAMDTYIDSYAQRLH------DEIDSRQ 205
             :| | |::   |  :   |  :      :::      |: |:  |      |:|::
Db     390 KDAQETARALESAAADYEEALTAVADGDLTRRVDASRDHDAMARIGHALNDMLDDIETSV 449

Qy     206 ELANAVATHVEAPLSSLEA-------TSQDVAERTDTMRARTDDQVDRMADVSREISSVS 258
           | | : ||     :||          ||:   | :     :| ||: :|: |:  :|
Db     450 AAATAFSDHVSDAAQRVEADAGDAIDAGTDVSTAVDEISDGATEQTDRLHEVAGEVDDLS 509

Qy     259 ASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAA 318
           || |||| |    :  |:  | :     |  | :||: || |: :  :     :: |    |
Db     510 ASAEEVAETVASLADTAGQAASAVDDGRQATEDAVETMDDVADDAEAAADAMDALDSEMA 569

Qy     319 DVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEE 378
           |:  :  || |||:|||||||||||||| |  |:|||||||||| ||||||: :  :|
Db     570 DIGEIVDVIADIADQTNMLALNASIEAARTGADGDGFAVVADEVKTLAEESRDAAEDIES 629

Qy     379 LVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDE 438
           :  :| :  :  |::   :  : :|   | :|   | ::   | |  :   ::  |||
Db     630 RLLALQGQVSDVADEMRATSDTVSDGRATVGDAATALDDVVSFVADTDTAAGQIRAATDR 689

Qy     439 QAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKLS 489
           || :   ||  || |   : | ||    :||:   |  |:  | :    |:
Db     690 QAHAASRVASAVDEVAGISQETAAQATAVADSAATQTDTLSSVDDAAADLA 740


RESULT 15
HTR1_HALSA
ID   HTR1_HALSA     STANDARD;      PRT;   535 AA.
AC   P33955;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sensory rhodopsin I transducer (HTR-I) (Methyl-accepting phototaxis
DE   protein I) (MPP-I).
GN   HTR1 OR HTRI OR HTR.
OS   Halobacterium salinarium.
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=L33;
RX   MEDLINE=90384855; PubMed=2205842;
RA   Ferrando E., Krah M., Marwan W., Oesterhelt D.;
RT   "A gene from S. pombe with homology to E. coli RNAse III blocks
RT   conjugation and sporulation when overexpressed in wild type cells.";
RL   Nucleic Acids Res. 18:5304-5304(1990).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=94244615; PubMed=8187768;
RA   Krah M., Marwan W., Vermeglio A., Oesterhelt D.;
RT   "Phototaxis of Halobacterium salinarium requires a signalling complex
RT   of sensory rhodopsin I and its methyl-accepting transducer HtrI.";
RL   EMBO J. 13:2150-2155(1994).
CC   -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC       SENSORY RHODOPSIN I (SR-I) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC       LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC   -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68591; CAA48578.1; -.
DR   PIR; S28466; S28466.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   Pfam; PF00015; MCPsignal; 1.
DR   Pfam; PF00672; HAMP; 1.
DR   SMART; SM00304; HAMP; 2.
DR   SMART; SM00283; MA; 1.
KW   Transducer; Photoreceptor; Transmembrane; Methylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1      8       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      9     29       POTENTIAL.
FT   DOMAIN       30     33       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     34     54       POTENTIAL.
FT   DOMAIN       55    535       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     265    265       METHYLATION (BY SIMILARITY).
FT   MOD_RES     272    272       METHYLATION (BY SIMILARITY).
FT   MOD_RES     279    279       METHYLATION (BY SIMILARITY).
FT   MOD_RES     463    463       METHYLATION (BY SIMILARITY).
FT   MOD_RES     472    472       METHYLATION (BY SIMILARITY).
```

```
SQ   SEQUENCE   535 AA;  56814 MW;  504A165A47FA8A45 CRC64;

  Query Match             18.8%;  Score 449.5;  DB 1;  Length 535;
  Best Local Similarity   27.7%;  Pred. No. 4.4e-13;
  Matches  141;  Conservative   86;  Mismatches  191;  Indels   91;  Gaps   15;

Qy      28 LDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTK 87
           ::  |   |:   |   : :|    || :     |    :|    :| |: :
Db      51 INAGETVGRIKEIGAQTERVANGNLEQEVTSTRTD----EFGSLADSIEQMRQSLRGRLN 106

Qy      88 TVEQ----LKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTG 143
            :|:    |:||| |          ::| | | |:  :                 |
Db     107 EMERTRSDLEETQTE----SETTREESEQAKQEAQSAE----------------REARE 145

Qy     144 LLDALADDVVADRGE--EAAAAVDELVARFLPMLKLLTFDQQI-------AMDTYIDSYA 194
           |     |    | ||  |::|  |              | |::       ||:|:  ::
Db     146 LAATYQDSSNRD-GETMESSATGD-------------FTQRVDVDTDPEAMETFGTAFN 190

Qy     195 QRLHDEIDSRQELANAVATHVEAPLSSLEATSQDV-AERTDTMRA------RTDDQVDRM 247
           | : |:: :       ||  :||    :  || |: |   ||: |      : :||   :
Db     191 QMM-DDLQATVRTVTTVADEIEAKTERMSETSADIEASAGDTVEAVSKIESQPNDQRTEL 249

Qy     248 ADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVT 307
              : ::  |||| ||:|:| ||:   |||    :     ::  ||  |: |:   |
Db     250 DSAADDVQQVSASAEEIAATIDDLASRSEDVATASDAARDSSKSALDEMSSIETEVDDPV 309

Qy     308 AGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAG--EAGEGFAVVADEVKAL 365
             || | :: |::  :  || || |||||||||||||||||    |:||:|||||| |
Db     310 GQVEHLRDQVAEITDIVDVITDIGEQTNMLALNASIEAARAGGNADGDGFSVVADEVKDL 369

Qy     366 AEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDA 425
           |||:::::  :  :||:: |:||:    :     |:  | | ||  :  :: | |:: :
Db     370 AEETQDRANEIAAVVEKVTAQTEDVTASIQHTRTRVESGSETVESTLLHIRTIADSISEV 429

Qy     426 ASGMQEVSTATDEQ-------AVSTEEVAEMVD-------------GVDDRAGEIAAAL 464
           :: : |:   | ||       | | | || : |             |  : | ||||:|
Db     430 SNSIDEIQRTTSEQAETVQSTATSVERVAGLSDDTTALASDAESPVIGQRESAEEIAASL 489

Qy     465 DDIADATDQQVRT-------VEEVRETVG 486
           :   :   :|:::         |  || |
Db     490 EQFQNTAVEQLQSRVASFTVATEDSETAG 518


Search completed: January  2, 2003, 12:41:59
Job time : 16 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:35:56 ; Search time 12 Seconds
(without alignments)
772.274 Million cell updates/sec

Title: US-09-455-978B-2
Perfect score: 2394
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVRTVEEVRETVGKLS 489

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	359.5	15.0	891	10	US-09-272-809-5	Sequence 5, Appli
2	292.5	12.2	531	1	US-08-976-063C-34	Sequence 34, Appl
3	254	10.6	433	8	US-08-945-038-6	Sequence 6, Appli
4	239.5	10.0	2310	9	US-09-991-496-120	Sequence 120, App
5	239.5	10.0	2310	10	US-09-874-923-120	Sequence 120, App
6	194	8.1	1786	9	US-09-742-096-3	Sequence 3, Appli
7	190	7.9	1879	9	US-09-971-536-70	Sequence 70, Appl
8	187	7.8	630	9	US-09-742-096-5	Sequence 5, Appli
9	177	7.4	1179	10	US-09-815-242-13608	Sequence 13608, A
10	176	7.4	540	9	US-09-742-096-22	Sequence 22, Appl
11	175.5	7.3	1289	9	US-09-712-363-259	Sequence 259, App
12	175.5	7.3	1798	10	US-09-845-583-8	Sequence 8, Appli
13	174	7.3	1179	10	US-09-815-242-13262	Sequence 13262, A
14	173.5	7.2	1798	10	US-09-938-275-9	Sequence 9, Appli
15	173	7.2	1786	10	US-09-873-676-113	Sequence 113, App
16	173	7.2	1786	10	US-09-938-275-6	Sequence 6, Appli
17	171.5	7.2	537	10	US-09-815-242-13463	Sequence 13463, A
18	171.5	7.2	1786	10	US-09-938-275-7	Sequence 7, Appli
19	171	7.1	2478	10	US-09-815-242-5816	Sequence 5816, Ap
20	171	7.1	2478	10	US-09-815-242-12967	Sequence 12967, A
21	169.5	7.1	1242	10	US-09-925-299-911	Sequence 911, App
22	169	7.1	1801	10	US-09-938-275-8	Sequence 8, Appli
23	168	7.0	1607	10	US-09-938-275-10	Sequence 10, Appl
24	167.5	7.0	2368	10	US-09-815-242-5635	Sequence 5635, Ap
25	167.5	7.0	2368	10	US-09-815-242-12389	Sequence 12389, A
26	164.5	6.9	1609	10	US-09-938-275-11	Sequence 11, Appl
27	163.5	6.8	1799	10	US-09-845-583-6	Sequence 6, Appli
28	163.5	6.8	2076	10	US-09-815-242-5815	Sequence 5815, Ap
29	163.5	6.8	2186	10	US-09-815-242-12913	Sequence 12913, A
30	161	6.7	1155	9	US-09-738-526-5764	Sequence 5764, Ap
31	159.5	6.7	600	9	US-09-738-526-5197	Sequence 5197, Ap
32	159	6.6	1958	12	US-10-028-946-4	Sequence 4, Appli
33	159	6.6	2025	10	US-09-815-242-5703	Sequence 5703, Ap
34	159	6.6	2053	9	US-10-017-216-2	Sequence 2, Appli
35	159	6.6	2054	12	US-10-028-946-2	Sequence 2, Appli
36	159	6.6	3158	10	US-09-815-242-12611	Sequence 12611, A
37	158	6.6	1162	10	US-09-815-242-11828	Sequence 11828, A
38	154	6.4	1597	9	US-10-017-216-6	Sequence 6, Appli
39	154	6.4	2055	9	US-10-017-216-4	Sequence 4, Appli
40	153.5	6.4	2437	10	US-09-815-242-5834	Sequence 5834, Ap
41	153.5	6.4	6281	10	US-09-815-242-12996	Sequence 12996, A
42	153	6.4	2835	10	US-09-885-535-4	Sequence 4, Appli
43	152	6.3	396	10	US-09-800-729-207	Sequence 207, App
44	150.5	6.3	463	9	US-09-738-526-3970	Sequence 3970, Ap
45	150.5	6.3	566	9	US-09-991-496-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-272-809-5
; Sequence 5, Application US/09272809
; Patent No. US20020022239A1
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C.
; TITLE OF INVENTION: Phytofiuors as fluorescent labels
; FILE REFERENCE: 2500.118US0
; CURRENT APPLICATION NUMBER: US/09/272,809
; CURRENT FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:cph5 locus SLL0041
; OTHER INFORMATION: (locus 1001300) an 891 aa protein,
; OTHER INFORMATION: methyl-accepting chemotaxis protein I. Homology
; OTHER INFORMATION: to tsr in last 250 aa.
US-09-272-809-5

Query Match 15.0%; Score 359.5; DB 10; Length 891;
Best Local Similarity 27.1%; Pred. No. 2.1e-14;
Matches 136; Conservative 86; Mismatches 197; Indels 83; Gaps 19;

```
Qy      23 ADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQP--LFEATADALVTDFYDHLESY----- 75
           |||:      | :|   |      : ::|   |  |    ||    | |  :| |
Db     418 ADRV------IVYRFDATWAGTVIVESVAEGYPKALGATIADPCFADSY--VEKYRSGRI 469

Qy      76 ERTQDLFANSTKT---VEQLK--ETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGL-- 128
           : |:|:: |:  |   : |||  | :|  :                | |    ::|||
Db     470 QATRDIY-NAGLTPCHIGQLKPFEVKANLV----------------APINYKGNLLGLLI 512

Qy     129 -----GP------DVYLGAYTRYYTGLL----DALADDVVADRGEEAAAAVDELVARFLP 173
                ||      :: |        ||     | ||   :|:   |     :::  | |
Db     513 AHQCSGPRDWHQNEIDLFGQLTVQVGLALERSDLLAQQKIAE--VEQRQMREKMQKRALE 570

Qy     174 MLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERT 233
           :  |:  |     | | ::    ||| :  :  ||    :   :: :: :   | |
Db     571 L--LMEVDPVSRGDLTIRAHV--TEDEIGTIADSYNATIESLRRIVTQVQTAASQFTETT 626
```

```
Qy    234 DT-------MRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGE 286
          ||       :  : : |   :|:    : ::: |:: ||  |       : |     |||
Db    627 DTNEVAVRQLACQANRQALDVAEALERLQAMNKSIQAVAENAAQAESAVQRATQTVDQGE 686

Qy    287 AAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
          | :  :  :  | |        |::||| :  :  |  :|   |:|||:|||||:||||
Db    687 DAMNRTVDGIVAIRETVAATAKQVKRLGESSQKISKVVNLIGSFADQTNLLALNAAIEAA 746

Qy    347 RAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVE 406
          ||| | ||||||||||::|| :| | :  : :||  :|||| | |: ::   :::  | :
Db    747 HAGEEGRGFAVVADEVRSLARQSAEATAEIAQLVATIQAETNEVVNAMEAGTEQVVVGTK 806

Qy    407 RVEEAMETLQEITDAVEDAASGMQE-VSTATDEQAVSTEE-------VAEMVD------- 451
          |||   :| :|| ||    ||: | :::|  ||: ::|        ||:: |
Db    807 LVEETRRSLNQIT-AVSAQISGLVEAITSAAIEQSQTSESVTQTMALVAQIADKNSSEAS 865

Qy    452 GVDDRAGEIAAALDDIADATDQ 473
          ||    |: |   : :|  |
Db    866 GVSATFKELLAVAQSLQEAVKQ 887


RESULT 2
US-08-976-063C-34
; Sequence 34, Application US/08976063C
; Publication No. US20020182697A1
;  GENERAL INFORMATION:
;    APPLICANT:  Alexander Steinbuchel; Horst Priefert; Jurgen Rabenhorst
;    TITLE OF INVENTION:  SYNTHETIC ENZYMES FOR THE PRODUCTION OF
;    TITLE OF INVENTION:  CONIFERYL ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN A
;    TITLE OF INVENTION:  ACID AND THEIR USE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG KRAMER SCHAEFER & BRISCOE
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 2.0 MB storage
;      COMPUTER:  HP VECTRA
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/976,063C
;      FILING DATE:  21-NOV-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  196 49 655.1 (Germany)
;      FILING DATE:  29-NOV-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 9998-CAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  531 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-976-063C-34

  Query Match           12.2%;  Score 292.5;  DB 1;  Length 531;
  Best Local Similarity  24.7%;  Pred. No. 1.1e-10;
  Matches  110;  Conservative   86;  Mismatches  192;  Indels   57;  Gaps   11;

Qy     63 ALVTDFYDHLESYERTQDL--FANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120
          | | |  |  :    :|  || |::       :| | : :|     :  || : : |

Db     90 AEVNDAIDIFDFQRGRSELLKFAASSR--------EASYSIEVG-----SNAAADKLQSG 136

Qy    121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEE-----------------AAA 162
          :  | | :  |  |        :  ::||   :: |: |:                  :|
Db    137 EPSDAL-MVADKKLNVEYEQLSSAVNALMGHLIEDQNEKVPLIYYMLGGVTLFTMLMSAY 195

Qy    163 AVDELVARFLPMLK-LLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSS 221
          :|  :  : :| ||  :   ::||     |    |  |||   |     :|  :   :
Db    196 SVWFISRQLVPPLKSTVQLAERIASGDLADVGDSRRKDEIGQLQSATRRMAIGLRNLVGD 255

Qy    222 L-EATSQDVAERTDTMRARTDDQVD------RMADVSREISSVSASVEEVASTADDVRRT 274
          : :: :| |:  :|         |:|       :| ||  :: :  :|: :| : ::
Db    256 IGQSRAQLVSSSSDLSAICAQAQIDVECQKLSVAQVSTAVNELVETVQAIAKSTEEAATV 315

Qy    275 SEDAEALAQQGEAAADDALATMTDIDEATDGVTA----GVEQLGERAADVESVTGVIDDI 330
          :  |:  |: ||:  : |:     |  |    |  |     :|:|   :| :  |  ||  :
Db    316 AVLADEKARGGESVVNKAV----DFIEHLSGDMAELGDAMERLQNDSAQINKVVDVIKAV 371

Qy    331 AEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEET 390
          |||||:|||||:|||||||| | ||||||||||:|||  ::: :  :| ||  :|  :|
Db    372 AEQTNLLALNAAIEAARAGEQGRGFAVVADEVRALAMRTQQSTKEIERLVVSLQQGSEAA 431

Qy    391 VDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTE------ 444
          : :     |  : |   ::|         ||    :  :::   ::|  :
Db    432 GELMRRGKVRTHDVVGLAQQAARRATRNYPAVAGIQAMNYQIAAGAEQQGAAVVQINQNM 491

Qy    445 -EVAEMVDGVDDRAGEIAAALDDIA 468
           || :| |    :||:   :  ::|
Db    492 LEVHKMADESAIKAGQTMKSSKELA 516


RESULT 3
US-08-945-038-6
; Sequence 6, Application US/08945038
; Patent No. US20020146423A1
;  GENERAL INFORMATION:
;    APPLICANT:  Doidge, Christopher V.
;    APPLICANT:  Lee, Adrian
;    APPLICANT:  Radcliff, Fiona J.
;    APPLICANT:  Hocking, Dianna M.
;    APPLICANT:  Webb, Elizabeth A.
;    TITLE OF INVENTION:  PROTECTIVE HELICOBACTER ANTIGENS
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY & LARDNER
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/945,038
;      FILING DATE:  23-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU96/00225
;      FILING DATE:  19-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PN2575
;      FILING DATE:  21-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PN3931
;      FILING DATE:  03-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PN7565
;      FILING DATE:  16-JAN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
```

```
;      REFERENCE/DOCKET NUMBER:  017227/0133
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 672-5300
;      TELEFAX:  (202) 672-5399
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  433 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-945-038-6

  Query Match             10.6%;  Score 254;  DB 8;  Length 433;
  Best Local Similarity   29.3%;  Pred. No. 1.7e-08;
  Matches   80;  Conservative   48;  Mismatches   97;  Indels   48;  Gaps    8;

Qy     191 DSYAQRLHDEIDSRQELANAVATHVE----------------APLSSLEATS---QDVAE 231
           ||     |  |::  | | | :  |                  | |::|| |   :: ||
Db      15 DSEIAELKKEVNLYQSLLN-LCLHEGFVGIKNNKVVFKSGNLASLNNLEEQSVHFKENAE 73

Qy     232 RTD------TMRARTDDQVDRMADVSRE-----------ISSVSASVEEVASTADDVRRT 274
             |      :::::  | |   :   :            :  ||::|    | :  :
Db      74 SVDLQGVSYSLKSQNIDGVQYFSLAKKTGCVGEYHKNDLFKTFCASLKEGLENAQESMQY 133

Qy     275 SEDAEAL----AQQGEAAADDALATMT----DID---EATDGVTAGVEQLGERAADVESV 323
                |    |: ||| : : | |:     ||:   |     |:  : | :|: ::  |
Db     134 FHQETGLLLNAAKNGEAHSTEGLGTVNKTGQDIESLYEKMQNATSLADSLNQRSNEITQV 193

Qy     324 TGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQM 383
           :|||||||||:|||||:|||||||| | |||||||||: |||:::: :  :  :|: |
Db     194 ISLIDDIAEQTNLLALNAAIEAARAGEHGRGFAVVADEVRKLAEKTQKATKEIVVVVKSM 253

Qy     384 QAETEETVDQLDEVNQRIGEGVERVEEAMETLQ 416
           | |  :      ::|  :      |||   |::
Db     254 QQEANDIQTNTHDINSIVSSIKGDVEELKSTVK 286


RESULT 4
US-09-991-496-120
; Sequence 120, Application US/09991496
; Patent No. US20020169285A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Coler, Rhea
;  APPLICANT:  Probst, Peter
;  APPLICANT:  Brannon, Mark
;  TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE: 210121.420C9
;  CURRENT APPLICATION NUMBER: US/09/991,496
;  CURRENT FILING DATE:  2001-11-20
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 120
;   LENGTH: 2310
;   TYPE: PRT
;   ORGANISM: Leishmania major and chagasi
US-09-991-496-120

  Query Match             10.0%;  Score 239.5;  DB 9;  Length 2310;
  Best Local Similarity   22.9%;  Pred. No. 8.1e-07;
  Matches  121;  Conservative   86;  Mismatches  224;  Indels   97;  Gaps   18;

Qy      14 RNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLE 73
           |   ||     ::  :  |:  ||       ||     ||    ||    | |
Db     960 RLAADGDEARQQLAANAEELQQRL-------DTATQQRAE---LEAQLARLAAD------ 1003

Qy      74 SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVY 133
           |  | | ||: :  ::|     :       | | : : |   |   ::   |
Db    1004 RDEARQQLAANAEELQQRLDTATQQ------RAELEAQVARLAANAEELQQRLDTATQ-- 1055

Qy     134 LGAYTRYYTGLLDALADDVVADRGE---EAAAAVDELVARFLPMLKLLTFDQQIAMDTYI 190
                   |:|    : ||| |   : ||  :||  |      | |  || |    :
Db    1056 -------QRAELEARVARLAADRDEARQQLAANAEELQQR------LDTATQQRA---EL 1099

Qy     191 DSYAQRL---HDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDT---MRARTDDQV 244
           ::   ||    :|:  | : |      :|| :: | | :::: :| ||    ||  : :|
Db    1100 EARVARLAANAEELQQRLDTATQQRAELEAQVARLAANAEELQQRLDTATQQRAELEARV 1159

Qy     245 DRMA---DVSREISSVSASVEEVASTAD----------DVRRTSEDAEALAQ------Q 284
           |:|   | :|:   ::|: ||:    |              | | : :|| | |      |
Db    1160 ARLAADRDEARQ--QLAANAEELQQRLDTATQQRAELEAQVARLAANAEELQQRLDTATQ 1217

Qy     285 GEAAADDALATM-TDIDEATDGVTAGVEQLGER----------------AADVESVTG 325
           |  :  || :  | |||   : |  |:| :|                   ||: | :
Db    1218 QRAELEAQLARLAADRDEARQQLAANAEELQQRLDTATQQRAELEAQVARLAANAEELQQ 1277

Qy     326 VIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQA 385
           :|   :|   |    :  ||   || :  |  |:|::   : : :|   :|  | :: |
Db    1278 RLDTATQQRAELEARVARLAADRDEARQQLAANAEELQQRLDTATQQRAELEARVARLAA 1337

Qy     386 ETEETVDQL----DEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAV 441
           : :|   ||    :|: ||:       ::  |   ::     :|    | : ||| ::|
Db    1338 DGDEARQQLAANAEELQQRLDTA---TQQRAELEAQVARLAANAEELQQRLDTATQQRAE 1394

Qy     442 STEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKLS 489
               || :    |:   ::||  :::    |   :   |:   | :|:
Db    1395 LEARVARLAADRDEARQQLAANAEELQQRLDTATQQRAELEAQVARLA 1442


RESULT 5
US-09-874-923-120
; Sequence 120, Application US/09874923
; Patent No. US20020081320A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Coler, Rhea
;  APPLICANT:  Probst, Peter
;  APPLICANT:  Brannon, Mark
;  TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE: 210121.420C8
;  CURRENT APPLICATION NUMBER: US/09/874,923
;  CURRENT FILING DATE:  2001-06-04
;  NUMBER OF SEQ ID NOS: 122
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 120
;   LENGTH: 2310
;   TYPE: PRT
;   ORGANISM: Leishmania major and chagasi
US-09-874-923-120

  Query Match             10.0%;  Score 239.5;  DB 10;  Length 2310;
  Best Local Similarity   22.9%;  Pred. No. 8.1e-07;
  Matches  121;  Conservative   86;  Mismatches  224;  Indels   97;  Gaps   18;

Qy      14 RNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLE 73
           |   ||     ::  :  |:  ||       ||     ||    ||    | |
Db     960 RLAADGDEARQQLAANAEELQQRL-------DTATQQRAE---LEAQLARLAAD------ 1003

Qy      74 SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVY 133
           |  | | ||: :  ::|     :       | | : : |   |   ::   |
Db    1004 RDEARQQLAANAEELQQRLDTATQQ------RAELEAQVARLAANAEELQQRLDTATQ-- 1055

Qy     134 LGAYTRYYTGLLDALADDVVADRGE---EAAAAVDELVARFLPMLKLLTFDQQIAMDTYI 190
                   |:|    : ||| |   : ||  :||  |      | |  || |    :
```

```
Db   1056 -------QRAELEARVARLAADRDEARQQLAANAEELQQR------LDTATQQRA---EL 1099

Qy    191 DSYAQRL---HDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDT---MRARTDDQV 244
          ::   ||    :|:  | : |      :|| :: | | :::: :| ||    ||  : :|
Db   1100 EARVARLAANAEELQQRLDTATQQRAELEAQVARLAANAEELQQRLDTATQQRAELEARV 1159

Qy    245 DRMA---DVSREISSVSASVEEVASTAD-----------DVRRTSEDAEALAQ------Q 284
           |:|   | :|:   ::|: ||:    |            | | : :|| | |      |
Db   1160 ARLAADRDEARQ--QLAANAEELQQRLDTATQQRAELEAQVARLAANAEELQQRLDTATQ 1217

Qy    285 GEAAADDALATM-TDIDEATDGVTAGVEQLGER-----------------AADVESVTG 325
            |  :  || :  | |||   : |  |:| :|                   ||: | :
Db   1218 QRAELEAQLARLAADRDEARQQLAANAEELQQRLDTATQQRAELEAQVARLAANAEELQQ 1277

Qy    326 VIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQA 385
           :|   :|      |     :  ||   || :  |  |:|::   : : :|   :|  | :: |
Db   1278 RLDTATQQRAELEARVARLAADRDEARQQLAANAEELQQRLDTATQQRAELEARVARLAA 1337

Qy    386 ETEETVDQL----DEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAV 441
           : :|   ||    :|: ||:       ::  |   ::     :|    | : ||| ::|
Db   1338 DGDEARQQLAANAEELQQRLDTA---TQQRAELEAQVARLAANAEELQQRLDTATQQRAE 1394

Qy    442 STEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKLS 489
             || :   |:   ::||  :::    |   :   |:   | :|:
Db   1395 LEARVARLAADRDEARQQLAANAEELQQRLDTATQQRAELEAQVARLA 1442


RESULT 6
US-09-742-096-3
; Sequence 3, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT:  DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 200773US0DIV
;  CURRENT APPLICATION NUMBER: US/09/742,096
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: US 08/973,642
;  PRIOR FILING DATE: 1998-02-06
;  PRIOR APPLICATION NUMBER: PCT/FR96/00894
;  PRIOR FILING DATE: 1996-06-12
;  PRIOR APPLICATION NUMBER: FR 95/07007
;  PRIOR FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: P. falciparum
US-09-742-096-3

  Query Match             8.1%;  Score 194;  DB 9;  Length 1786;
  Best Local Similarity   24.5%;  Pred. No. 0.00031;
  Matches  100;  Conservative   85;  Mismatches  139;  Indels   84;  Gaps   22;

Qy    146 DALADDV---VADRGEEAAAA-VDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEI 201
          :::|::|   ||:  ||: |  |:|:||  :  :   | :: :|  : ::| |  : : :
Db    565 ESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVA-PSVVESVAPSVEESV 623

Qy    202 DSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREI------S 255
          :   |   :|| :||   |  |   : ||| : : | | ::: :|    ||
Db    624 EENVE--ESVAENVEE--SVAENVEESVAENVEEIVAPTVEEI--VAPTVEEIVAPSVVE 677

Qy    256 SVSASVEEVA-----STADDVRRT-SEDAEALAQQGEAAADDALATMTD-IDEATDGVT 307
          ||: ||||       | |::|  : :|: |      |: |::   :: : ::|    |
Db    678 SVAPSVEESVEENVEESVAENVEESVAENVE------ESVAENVEESVAENVEEI---VA 728

Qy    308 AGVEQLGERAAD-------VESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFA---- 356
            ||::     :       ||||   :::  |:    ::  ::| : |  |  |  |
Db    729 PTVEEIVAPTVEEIVAPSVVESVAPSVEESVEENVEESVAENVEESVAENVEESVAENVE 788

Qy    357 -VVADEVKALAEESREQST--RVEELVEQMQA--------------ETEETVD----QL 394
```

```
          ||  |: :   | |:|    ||| | :  |               ||||  |    ::
Db    789 ESVAPTVEEIVAPSVEESVAPSVEESVAENVATNLSDNLLSNLLGGIETEEIKDSILNEI 848

Qy    395 DEVNQRIGEGV-ERVEEA----------METLQEIT---DAVEDAASGMQE-VSTATDE 438
          :|| : :   : | |||             :| :|| |   | :|:    : | | :|  |
Db    849 EEVKENVVTTILENVEETTAESVTTFSNILEEIQENTITNDTIEEKLEELHENVLSAALE 908

Qy    439 QAVSTEEVAEMVDGVDDRAGEIAAAL-DDIADATDQQVRTVEEVRETV 485
            | ||  |::| :::   |:|  | : :  | ::   |: |: | :
Db    909 NTQSEEEKKEVIDVIEEVKEEVATTLIETVEQAEEKSANTITEIFENL 956


RESULT 7
US-09-971-536-70
; Sequence 70, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka
;  APPLICANT:  Bloksberg, Leonard
;  APPLICANT:  Lubbers, Mark
;  APPLICANT:  Dekker, James
;  APPLICANT:  Christensson, Anna
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul
;  APPLICANT:  Reid, Julian
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth
;  TITLE OF INVENTION:  Using Them
;  FILE REFERENCE: 1043c2
;  CURRENT APPLICATION NUMBER: US/09/971,536
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
;  PRIOR FILING DATE: 2000-08-08
;  PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: PCT/NZ01/00160
;  PRIOR FILING DATE: 2001-08-08
;  NUMBER OF SEQ ID NOS: 83
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70
;   LENGTH: 1879
;   TYPE: PRT
;   ORGANISM: Lactobacillus rhamnosus
US-09-971-536-70

  Query Match             7.9%;  Score 190;  DB 9;  Length 1879;
  Best Local Similarity   21.4%;  Pred. No. 0.00056;
  Matches  119;  Conservative   83;  Mismatches  185;  Indels  168;  Gaps   25;

Qy     12 DVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEATADA-------- 63
          | :  ||  |          |::        || |     | ::   :|  ||
Db    992 DAKQAIDAEA---------AKVT-----KAIDQDPTLTAAEKKAQKQAVTDAETKAKAAI 1037

Qy     64 ---LVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQR---- 116
             || |  |           |:  ||::      |:|  |:     | : || :
Db   1038 DATLVADAID---------QALADGIKTID------AQYQTGIA---LDKQKAAAKQTID 1079

Qy    117 ARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLK 176
          |   |: :   :  || | |  :       |     |||   :| ||:|:
Db   1080 AEAAKVSE--AIDQDVTLTADQK-------ATQKQAVADEATKAKAAIDQ---------- 1120

Qy    177 LLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTD-- 234
          |     :   ||       : ||:: : | |:     :    :::| :  |::  |
Db   1121 --ASDADAVIQATIDGI-----EAIDAQHQSATALDKQKQQAKQAIDAEAAKVSKAIDQD 1173

Qy    235 -----TMRARTDDQVDRMADVSRE-----------------ISSVSASVEEVASTADDV 271
                | :|     |   || :::                 | :: |  :   : ||
Db   1174 VTLTATQKADQKQAVIAEADKAKKLIDAAGNADGIKQAESDGIKAIDAQHQSSQALADRK 1233

Qy    272 R--RTSEDAEAL----------------AQQGEAAADDALATMTDIDEA---------- 302
          |  :|: ||||                | | :|  |:|      || |
Db   1234 RDAKTAIDAEAAKETAAIDHDATLTANEKASQKQAVTDEATKAKKAIDAAKQADAVDQAK 1293
```

```
Qy    303 TDGVTA------GVEQLGERAADVESVTGVID-DIAEQTNMLALNASIEAARAGEAGEGF 355
          |||: |        : | :| || :    ||| : |: | :  : :: |:     :|
Db   1294 TDGIKAIDAQHHSGQALDDRKADAKQ---VIDAEAAKVTAAIDQDNTLTKAQKAAQKQGV 1350

Qy    356 AVVADEVKALAEESREQSTRVEELVEQMQA-----ETEETVD-QLDEVNQRIGEGVERVE 409
          |  ||: |   : : :     :     :||     :  :|:| : |:  | | |   :|
Db   1351 ATEADKAKQAIDAAGDADAVDQAKTAGIQAIDAQHKAGKTIDSRHDDAKQAIDEEAAKVI 1410

Qy    410 EAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIAD 469
          :|::    :|.| ::|           :|||:||  |:      | ||:        ||
Db   1411 KAIDQDPTLTAAQKEA----------QKQAVATE--ADKAKKAIDAAGD--------AD 1449

Qy    470 ATDQQ----VRTVEE 480
          | ||     :: ::|
Db   1450 AVDQAKTAGIKAIDE 1464


RESULT 8
US-09-742-096-5
; Sequence 5, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT:  DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 200773US0DIV
;  CURRENT APPLICATION NUMBER: US/09/742,096
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: US 08/973,642
;  PRIOR FILING DATE: 1998-02-06
;  PRIOR APPLICATION NUMBER: PCT/FR96/00894
;  PRIOR FILING DATE: 1996-06-12
;  PRIOR APPLICATION NUMBER: FR 95/07007
;  PRIOR FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 630
;   TYPE: PRT
;   ORGANISM: P. falciparum
US-09-742-096-5

  Query Match             7.8%;  Score 187;  DB 9;  Length 630;
  Best Local Similarity   21.5%;  Pred. No. 0.00025;
  Matches   85;  Conservative   87;  Mismatches  138;  Indels   86;  Gaps   16;

Qy    150 DDVVADRGEEAAA-AVDELVARFLPMLK--LLTFDQQIAMDTYIDSYAQRLHDEIDSRQE 206
          :::||   ||: |  |:|:||   | ::  :    ::| : :  :| |  : : :   :
Db    263 EEIVAPTVEESVAPTVEEIVA---PTVEESVAPTVEEIVVPSVEESVAPSVEESV--AEN 317

Qy    207 LANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVA- 265
          :  :|| :||   |  |   : |||  :   |   :::           |: ||||:
Db    318 VEESVAENVEE--SVAENVEESVAENVEESVAENVEEI-----------VAPSVEEIVA 363

Qy    266 -----STADDVRRTSEDAEALAQ-----QGEAAADDALATMTDIDEATDGVTAGVEQLGE 315
               | |::|    |   |:        :  |   |  : : :  |  :|:: |
Db    364 PTVEESVAENVATNLSD-NLLSNLLGGIETEEIKDSILNEIEEVKE--NVVTTILEKVEE 420

Qy    316 RAAD-VESVTGVIDDIAEQT------------NMLALNASIEAARAGEAGEGFAVVADE 361
            |: | : : ::::| | |             :   |:|::|  :: |  :    | :|
Db    421 TTAESVTTFSNILEEIQENTITNDTIEEKLEELHENVLSAALENTQSEEEKKEVIDVIEE 480

Qy    362 VKALAEESREQSTRVEELVEQMQAETEETVDQL--------DEVNQRIGEGVERVEEAM- 412
          ||       | :| : | ||| : |:| |: ::          | |::: | :|:: | :
Db    481 VK------EEVATTLIETVEQAEEESESTITEIFENLEENAVESNEKVAENLEKLNETVF 534

Qy    413 -ETLQEITDAVE-------------------DAASGMQEVSTATDEQAVSTEEVAEMVD 451
           | :: : ||                     |   |:||      ::: |  | :  :
Db    535 NTVLDKVEETVEISGESLENNEMDKAFFSEIFDNVKGIQENLLTGMFRSIETSIVIQSEE 594

Qy    452 GVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGK 487
           ||     ::: ||:| :  :  :  :| :  | |:
```

```
Db    595 KVDLNENVVSSILDNIENMKEGLLNKLENISSTEGE 630


RESULT 9
US-09-815-242-13608
; Sequence 13608, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13608
;   LENGTH: 1179
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-815-242-13608

  Query Match             7.4%;  Score 177;  DB 10;  Length 1179;
  Best Local Similarity   18.8%;  Pred. No. 0.002;
  Matches  126;  Conservative  114;  Mismatches  206;  Indels  224;  Gaps   28;

Qy     30 EAEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLES-YERTQDLFANSTKT 88
          |||:   |:|:  | | |  |  |: :     :| |::    :|:  | :: |
Db    364 EAEL---LAFSD-DPDQMIELLRERFVALLQEEADVSNQLTRIENELENSRQLSQKQADQ 419

Qy     89 VEQLKE----------TQAEYL----------------LGLGRGEYDTEYAAQ------- 115
          :|:|||           | | |                :   : |  | | ||
Db    420 LEKLKEQLAIAKEKASQQKEELETAKEQVQKLLADYQAIAKEQEEQKTSYQAQQSQLFDR 479

Qy    116 -------RARIGKIHDVLGLGPDVYLGAYT--------------------RYYTGL- 144
                 :||   : ::|    : | |  :                        | | |
Db    480 LDNLKNKQARAQSLENILRNHSNFYAGVKSVLQEKDRLGGIIGAVSEHLTFDMHYQTALE 539

Qy    145 --LDALADDVVADRGEEAAAAVDELV------ARFLPM--LKLLTF------------- 180
            | | :  :: :    |  |:| |        |||:  :|  |
Db    540 IALGASSQHIIVEDENAATKAIDFLKRNRAGRATFLPLTTIKARTISSQNQDAIAVSPGF 599

Qy    181 ----DQQIAMDTYIDSYAQRL------HDEIDSRQELANAVATHVEAPLSSLEATSQDVA 230
              |: :  || :::  : |       | ::  :| |  |   |   : :|: |
Db    600 LGMADELVTFDTRLEAIFKNLLATTAIFDTVEEAREAARQVRYQVR--MVTLDGTEL--- 654

Qy    231 ERTDTMRA----RTDDQV---DRMADVSREISSVSASV--EEVA--STADDVRRTSEDAE 279
           ||    |    | :: :    :  : :|||:  ||:  |||| :  |:: | :|  |
Db    655 -RTGGSYAGGANRQNNSIFIKPELEQLQKEIAEEEASLDSEEVALKTLQDEMARLTESLE 713

Qy    280 ALAQQGEAA---------------ADDALATMTDI-DEATDGVTAG---------VEQL 313
          |: |||  |                  : | |   : :|  | :: |          |:|
```

```
Db    714 AIKSQGEQARIQEQGLFLAYQQTSQQVEELETLWKLQEEEIDRLSEGDWQADKENCQERL 773

Qy    314 GERAADVESVTGVIDDIAEQTNMLA---LNASIEAARA--------------------G 349
          |:| :::  |::|    | :     |   | |:|                        |
Db    774 AAIASDKQNLEAEIEEIKSNKNAIQERYQNLQEELAQARLLKTELQGQKRYEVADIERLG 833

Qy    350 EAGEGFAVVADEVKAL-------------------AEESREQSTRVEELVEQMQAETEET 390
          :  :   :  :|:: :                    |||:: | | ::: : : | | ::
Db    834 KELDNLNIEQEEIQRMLQEKVDNLEKVDTDLLNQQAEEAKTQKTNLQQGLIRKQFELDDI 893

Qy    391 VDQLDEVNQRIGEG-----------------VERVEEAMETLQ-EITDAVEDAASGMQEV 432
            |||::   : :                   |:| | :  || ::||  : : :   |
Db    894 EGQLDDIASHLDQARQQNEEWIRKQTRAEAKKEKVSERLRHLQSQLTDQYQISYTEALEK 953

Qy    433 STATDEQAVSTEEVAEM-----------VDGVD------DRAGEIAAALDDIADATDQQV 475
          :   :   :: :|| ::           :: :|      :|   : :  |||  | :  :
Db    954 AHELENLNLAEQEVKDLEKAIRSLGPVNIEAIDQYEEVYNRLDFLNSQRDDILSAKNLLL 1013

Qy    476 RTVEEVRETV 485
            |: |: : |
Db   1014 ETITEMNDEV 1023


RESULT 10
US-09-742-096-22
; Sequence 22, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT:  DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 200773US0DIV
;  CURRENT APPLICATION NUMBER: US/09/742,096
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: US 08/973,642
;  PRIOR FILING DATE: 1998-02-06
;  PRIOR APPLICATION NUMBER: PCT/FR96/00894
;  PRIOR FILING DATE: 1996-06-12
;  PRIOR APPLICATION NUMBER: FR 95/07007
;  PRIOR FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
;   LENGTH: 540
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
US-09-742-096-22

  Query Match             7.4%;  Score 176;  DB 9;  Length 540;
  Best Local Similarity   25.7%;  Pred. No. 0.00095;
  Matches   96;  Conservative   78;  Mismatches  142;  Indels   58;  Gaps   20;

Qy    146 DALADDV---VADRGEEAAAA-VDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEI 201
          :::|::|   ||:  ||: |  |:|:||  :  :    | :: :|  : ::| |  : : :
Db     39 ESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVA-PSVVESVAPSVEESV 97

Qy    202 DSRQELANAVATHVEAPL------SSLEATSQDVAERTDTMRARTDDQVDRMADVSREIS 255
          :   |   :|| :|  :       |  |   : |||  :   |   :::  :|    ||
Db     98 EENVE--ESVAENVEESVAENVEESVAENVEESVAENVEESVAENVEEI--VAPTVEEI- 152

Qy    256 SVSASVEE------VASTADDVRRTSED--AEALAQQGEAA----ADDALATMTDIDEAT 303
           |: :|||      | | |  |  : |:   |::|:  | :      ::::|   :    |
Db    153 -VAPTVEEIVAPSVVESVAPSVEESVEENVEESVAENVEESVAENVEESVAENVEESVAE 211

Qy    304 DGVTAGVEQLGERAAD--VESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADE 361
          :   :  | : |  |:   |||   :::|    |    :  ::|   |    |    ||
Db    212 NVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVES---VAPS 268

Qy    362 VKALAEESREQSTRVEELVEQMQAE-TEETVDQLDEVNQRIGEGVERV------EEAMET 414
          |:   ||: |:|  | | ||:  |  ||:|   : | : : | || :      |    |
Db    269 VEESVEENVEES--VAENVEESVAENVEESV--AENVEESVAENVEEIVAPTVEEIVAPT 324
```

```
Qy    415 LQEIT--DAVEDAASGMQEVSTATDEQAVS---TEEVAEMVDGVDDRAGEIAAALDDIAD 469
          ::||     ||  |  ::|      |::|:    | ||| |:        :|  :::|
Db    325 VEEIVAPSVVESVAPSVEESVEENVEESVAENVEESVAENVE------ESVAENVEEIVA 378

Qy    470 ATDQQV--RTVEEV 481
            | :::   ||||:
Db    379 PTVEEIVAPTVEEI 392


RESULT 11
US-09-712-363-259
; Sequence 259, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
;  APPLICANT: Eisenberg, David
;  APPLICANT:  Rotstein, Sergio H.
;  APPLICANT:  Marcotte, Edward M.
;  TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;  TITLE OF INVENTION:  INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
;  FILE REFERENCE: 07419-032001
;  CURRENT APPLICATION NUMBER: US/09/712,363
;  CURRENT FILING DATE:  2000-11-13
;  PRIOR APPLICATION NUMBER: PCT/US00/02246
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/179,531
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: 60/117,844
;  PRIOR FILING DATE: 1999-01-29
;  PRIOR APPLICATION NUMBER: 60/118,206,
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: 60/126,593
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/134,093
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/134,092
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/165,124
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/165,086
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 292
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 259
;   LENGTH: 1289
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
US-09-712-363-259

  Query Match             7.3%;  Score 175.5;  DB 9;  Length 1289;
  Best Local Similarity   24.0%;  Pred. No. 0.0027;
  Matches  134;  Conservative   62;  Mismatches  197;  Indels  165;  Gaps   24;

Qy     25 RIG-LDEAEIAWRLSFTGIDDD---TMAALAAEQPLFEATADALVTDFYDHLESYERTQD 80
          ||| ||: |:       |:|:    |:|||          ||  | :    |:| ||
Db    524 RIGELDQGEV-------GLDEHHERTVAAL--------RLADERVAE----LQSAERAA- 563

Qy     81 LFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLG----A 136
                                    : : |: ||||    | | :|     |    |
Db    564 --------------------------ERQVASLRARI----DALAVGLQRKDGAAWLA 591

Qy    137 YTRYYTGLLDALADDVVADRGEEA-----------AAAVDELVARFLPMLKLLTFD---- 181
          : |   ||  ::|  |     | ||            | ||| | |    :  |   |
Db    592 HNRSGAGLFGSIAQLVKVRSGYEAALAAALGPAADALAVDGLTAAGSAVSALKQADGGRA 651

Qy    182 -----------QQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDV 229
                      | |    : | ||   | ::|  :|  |:     | || : |   |:
Db    652 VLVLSDWPAPQAPQSASGEMLPSGAQWALDLVESPPQLVGAMI----AMLSGV-AVVNDL 706

Qy    230 AERTDTMRARTDDQVDRMADVSREI---------SSVSASVEEVASTADDVRRTSEDAEA 280
          |    : | :    |   |  ::          |    |  || | |  |     |||
Db    707 TEAMGLVEIRPE---LRAVTVDGDLVGAGWVSGGSDRKLSTLEVTSEIDKARSELAAAEA 763
```

```
Qy    281 LAQQGEA-------------AADDALATMTDIDEATDGVTAGVEQLGE-----RAADVE 321
          || |  |              ||: ||| : : | |   ::|  ||||      |||: |
Db    764 LAAQLNAALAGALTEQSARQDAAEQALAALNESDTA---ISAMYEQLGRLGQEARAAEEE 820

Qy    322 ------------SVTGVIDDI---------AEQTNML---------ALNASIEAARAGE 350
                      |  :||:          |::|  :          |::|: : ||  |
Db    821 WNRLLQQRTEQEAVRTQTLDDVIQLETQLRKAQETQRVQVAQPIDRQAISAAADRARGVE 880

Qy    351 AGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEE 410
            ||    | :| |   |  | |     |:   |       |                | :
Db    881 VEARLAVRTAEERANAVRGRADSLRRAAAAER---EARVRAQQARAARLHAAAVAAAVAD 937

Qy    411 AMETLQEITDAVEDAASGMQEVSTATDEQAVST-EEVAEMVDGVDDRAGEIAAAL--DDI 467
              |         | || ::: | |  :| ::    | : |: :  | ||:  :|  |::
Db    938 CGRLLAGRLHRAVDGASQLRDASAAQRQQRLAAMAAVRDEVNTLSARVGELTDSLHRDEL 997

Qy    468 ADATDQQVRTVEEVRETV 485
          |:|  |    :|:: : |
Db    998 ANA--QAALRIEQLEQMV 1013


RESULT 12
US-09-845-583-8
; Sequence 8, Application US/09845583
; Patent No. US20020142954A1
; GENERAL INFORMATION:
;  APPLICANT: Burgeson, Robert
;  APPLICANT:  Brunken, William Joseph
;  APPLICANT:  Champliaud, Marie-France
;  APPLICANT:  Hunter, Dale
;  TITLE OF INVENTION: LAMININ 15 AND USES THEREOF
;  FILE REFERENCE: 10287-056001
;  CURRENT APPLICATION NUMBER: US/09/845,583
;  CURRENT FILING DATE:  2001-04-30
;  PRIOR APPLICATION NUMBER: US 60/200,863
;  PRIOR FILING DATE: 2000-05-01
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 1798
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-845-583-8

  Query Match             7.3%;  Score 175.5;  DB 10;  Length 1798;
  Best Local Similarity   20.4%;  Pred. No. 0.0039;
  Matches  112;  Conservative   80;  Mismatches  230;  Indels  127;  Gaps   17;

Qy      7 TLVTADVRNGID-----GHALADRIGLDEAEIAWRLSFTGID--------DDTMAALAAE 53
          | : ||: :  |      |||:   ||:   :|  |:   :|         : : |  :
Db   1277 TQLEADLTDVQDENFNANHALS---GLERDRLALNLTLRQLDQHLDLLKHSNFLGAYDSI 1333

Qy     54 QPLFEATADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYA 113
          :     :|:|          |    |  |   |  :          : | |:    : ::::::
Db   1334 RHAHSQSAEA---------ERRANTSALAVPSPVSNSASARHRTEALMDAQKEDFNSKHM 1384

Qy    114 AQRARIGKI----------------------HDVL----GLG----------------- 129
          | :  :||:                       ||      | |
Db   1385 ANQRALGKLSAHTHTLSLTDINELVCGAQGLHHDRTSPCGGAGCRDEDGQPRCGGLSCNG 1444

Qy    130 ----PDVYLGAYTRYYTGLLDALAD-----DVVADRGEEAAAAVDELVARFLPMLKLLTF 180
              |: ||        |  |||:       ||:   :|: |     |
Db   1445 AAATADLALGRARHTQAELQRALAEGGSILSRVAETRRQASEAQQRAQA---------AL 1495

Qy    181 DQQIAMDTYIDSYAQRLHDEIDSRQELAN----------AVATHVEAPLSSLEATSQDVA 230
          |:  |    ::   | | : | | ::  |          |||  |      || :    |
Db   1496 DKANASRGQVEQANQELQELIQSVKDFLNQEGADPDSIEMVATRV------LELSIPASA 1549

Qy    231 ERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRT---SEDAEALAQQGEA 287
          |:   :     ::|  :|||  ::      |        | ||    :|| :  |:  :|
Db   1550 EQIQHLAGAIAERVRSLADVDAILARTVGDVRRAEQLLQDARRARSWAEDEKQKAETVQA 1609

Qy    288 AADDALATMTDIDEATDGVTAGVE-------QLGERAADVESVTGVIDDIAEQTNMLALN 340
```

```
          | ::|        |  |  |          |: || |  |       : | |     |:
Db   1610 ALEEAQRAQGIAQGAIRGAVADTRDTEQTLYQVQERMAGAERALSSAGERARQ-----LD 1664

Qy    341 ASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQR 400
          | :|| :   ||   |    |  | : : | |    |:|:     :  :||  | |   |
Db   1665 ALLEALKLKRAGNSLAASTAEETAGSAQGRAQ--EAEQLLRGPLGDQYQTVKALAE---R 1719

Qy    401 IGEGV----ERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR 456
            :||     | |:  :  :::  | :|    :||:    :|   : |  | :||:: |
Db   1720 KAQGVLAAQARAEQLPDEARDLLQAAQDKLQRLQELEGTYEENERALESKAAQLDGLEAR 1779

Qy    457 AGEIAAALD 465
            :  |::
Db   1780 MRSVLQAIN 1788


RESULT 13
US-09-815-242-13262
; Sequence 13262, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13262
;   LENGTH: 1179
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-815-242-13262

  Query Match             7.3%;  Score 174;  DB 10;  Length 1179;
  Best Local Similarity   19.2%;  Pred. No. 0.003;
  Matches  113;  Conservative  101;  Mismatches  201;  Indels  174;  Gaps   24;

Qy     49 ALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEY 108
          |:| ||   : :  |  :  :| |:| :  |       :::| :   : :  |:
Db    457 AIAKEQEEQKTSYQAQQSQLFDRLDSLKNKQ---ARAQSLENILRNHSNFYAGV----- 507

Qy    109 DTEYAAQRARIGKIHDVLG--LGPDVYLGAYTRYYTGL---LDALADDVVADRGEEAAAA 163
              :: |:| |   :   |  |||      | | |   | | : :: :  | |  |
Db    508 -KSVLQEKDRLGGIIGAVSEHLTFDVY-----YQTALEIALGASSQHIIVEDEESATKA 560

Qy    164 VDELV------ARFLPM--LKLLTF----------DQQIAMDTYIDSYAQRL 197
          :| |       | |||:  :|  |                |: :  || :::  : |
Db    561 IDFLKRNRVGRATFLPLTTIKARTISSQNQDAIAVSPGFLGMADELVTFDTRLEAIFKNL 620

Qy    198 ------HDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRA----RTDDQV--- 244
```

```
Db    621 LATTAIFDTVEEAREAARQVRYQVR--MVTLDGTEL----RTGGSYAGGANRQNNSIFIK 674

Qy    245 DRMADVSREISSVSASV--EEVA--STADDVRRTSEDAEALAQQGEAA---------- 288

Db    675 PELEQLQKEIAADEASLGSEEAALKTLQDQMAALTERLEAIKSQGEQARIQEQGLSLAYQ 734

Qy    289 ----ADDALATMTDI-DEATDGVTAG---------VEQLGERAADVESVTGVIDDIAEQT 334

Db    735 QTSQQVEELETLWKLQEEEIDRLSEGDWQADKEKCQESLATIASEKQNLEAEIEEIKSNK 794

Qy    335 NMLA---LNASIEAARA--------------------GEAGEGFAVVADEVKAL----- 365

Db    795 NAIQERYQNLQEEVAQARLLKTELQGQKRYEVADIERLGKELDNLNIEQEEIQRMLQEKV 854

Qy    366 -------------AEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEG------- 404

Db    855 DNLEKVDTELLSQQAEESKTQKTNLQQGLIRKQFELDDIEGQLDDIASHLDQARQQNEEW 914

Qy    405 ----------VERVEEAMETLQ---------EITDAVED---------AASGMQEVSTAT 436

Db    915 IRKQTRAEAKKEKVSERLRHLQNQLTDQYQISYTEALEKAHELENLNLAEQEVQDLEKAI 974

Qy    437 DEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETV 485

Db    975 RSLGPVNLEAIDQYEEVHNRLDFLNSQRDDILSAKNLLLETITEMNDEV 1023


RESULT 14
US-09-938-275-9
; Sequence 9, Application US/09938275
; Patent No. US20020111309A1
; GENERAL INFORMATION:
;  APPLICANT: Gerardo Castillo
;  APPLICANT:  Alan Snow
;  TITLE OF INVENTION: Therapeutic and Diagnostic Applications
;  TITLE OF INVENTION:  of Laminin and Laminin-Derived Protein Fragments
;  FILE REFERENCE: PROTEO.P03
;  CURRENT APPLICATION NUMBER: US/09/938,275
;  CURRENT FILING DATE:  2001-08-16
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 1798
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: Swissprot P55268
;   DATABASE ENTRY DATE: 1996-10-01
US-09-938-275-9

  Query Match             7.2%;  Score 173.5;  DB 10;  Length 1798;
  Best Local Similarity   20.2%;  Pred. No. 0.0051;
  Matches  111;  Conservative   83;  Mismatches  228;  Indels  127;  Gaps   17;

Qy      7 TLVTADVRNGID-----GHALADRIGLDEAEIAWRLSFTGID--------DDTMAALAAE 53

Db   1277 TQLEADLTDVQDENFNANHALS---GLERDRLALNLTLRQLDQHLDLLKHSNFLGAYDSI 1333

Qy     54 QPLFEATADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYA 113

Db   1334 RHAHSQSAEA---------ERRANTSALAVPSPVSNSASARHRTEALMDAQKEDFNSKHM 1384

Qy    114 AQRARIGK------------IHDVL---------------GLG----------------- 129

Db   1385 ANQRALGKLSAHTHTLSLTDINELVCGAPGDAPCATSPCGGAGCRDEDGQPRCGGLSCNG 1444

Qy    130 ----PDVYLGAYTRYYTGLLDALAD-----DVVADRGEEAAAAVDELVARFLPMLKLLTF 180

Db   1445 AAATADLALGRARHTQAELQRALAEGGSILSRVAETRRQASEAQQRAQA---------AL 1495

Qy    181 DQQIAMDTYIDSYAQRLHDEIDSRQELAN----------AVATHVEAPLSSLEATSQDVA 230

Db   1496 DKANASRGQVEQANQELQELIQSVKDFLNQEGADPDSIEMVATRV------LELSIPASA 1549

Qy    231 ERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRT---SEDAEALAQQGEA 287

Db   1550 EQIQHLAGAIAERVRSLADVDAILARTVGDVRRAEQLLQDARRARSWAEDEKQKAETVQA 1609

Qy    288 AADDALATMTDIDEATDGVTAGVE-------QLGERAADVESVTGVIDDIAEQTNMLALN 340

Db   1610 ALEEAQRAQGIAQGAIRGAVADTRDTEQTLYQVQERMAGAERALSSAGERARQ-----LD 1664

Qy    341 ASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQR 400

Db   1665 ALLEALKLKRAGNSLAASTAEETAGSAQGRAQ--EAEQLLRGPLGDQYQTVKALAE---R 1719

Qy    401 IGEGV----ERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR 456

Db   1720 KAQGVLAAQARAEQLRDEARDLLQAAQDKLQRLQELEGTYEENERALESKAAQLDGLEAR 1779

Qy    457 AGEIAAALD 465

Db   1780 MRSVLQAIN 1788


RESULT 15
US-09-873-676-113
; Sequence 113, Application US/09873676
; Patent No. US20020077289A1
; GENERAL INFORMATION:
;  APPLICANT: MacDonald, Nicholas J.
;  APPLICANT:  Sim, Kim L.
;  TITLE OF INVENTION: Angiostatin and Endostatin Binding Proteins and Methods of Use
;  FILE REFERENCE: 05213-0378 (43170-259333)
;  CURRENT APPLICATION NUMBER: US/09/873,676
;  CURRENT FILING DATE:  2001-06-04
;  PRIOR APPLICATION NUMBER: US 60/209,065
;  PRIOR FILING DATE: 2000-06-02
;  PRIOR APPLICATION NUMBER: US 60/289,387
;  PRIOR FILING DATE: 2001-05-08
;  NUMBER OF SEQ ID NOS: 123
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-873-676-113

  Query Match             7.2%;  Score 173;  DB 10;  Length 1786;
  Best Local Similarity   19.7%;  Pred. No. 0.0055;
  Matches  103;  Conservative   92;  Mismatches  192;  Indels  136;  Gaps   20;

Qy     45 DTMAALAAEQP------LFEATADALVTDFYDHLESYE-RTQDLFANSTKTVEQLK--ET 95

Db   1232 DILAQSPAAEPLKNIGNLFE-EAEKLIKDVTEMMAQVEVKLSDTTSQSNSTAKELDSLQT 1290

Qy     96 QAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVAD 155

Db   1291 EAESL------DNTVKELAEQLEFIKNSDIRGA-----LDSITKYFQMSLEA-------- 1331

Qy    156 RGEEAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHV 215

Db   1332 -EERVNASTTEPNS---------TVEQSALMRDRVEDVMMERESQFKEKQEEQARLLDEL 1381

Qy    216 EAPLSSLEATSQDVAERT--------------DTMRARTDD------------------- 242

Db   1382 AGKLQSLDLSA--AAEMTCGTPPGASCSETECGGPNCRTDEGERKCGGPGCGGLVTVAHN 1439

Qy    243 QVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDAL----ATMTD 298

Db   1440 AWQKAMDLDQDVLSALAEVEQLS-------KMVSEAKLRADEAKQSAEDILLKTNATKEK 1492

Qy    299 IDEATDGVTAGVEQ----LGERAADVESVTGV------------------IDDIAEQTN 335

Db   1493 MDKSNEELRNLIKQIRNFLTQDSADLDSIEAVANEVLKMEMPSTPQQLQNLTEDIRERVE 1552
```

```
Qy     336 MLALNASI----------------EAARAGEAGEGFAVVADEVKALAEESREQSTRVEEL 379
           |:    |                 || || ::     | || ||   ||: :     |:
Db    1553 SLSQVEVILQHSAADIARAEMLLEEAKRASKSATDVKVTADMVKEALEEAEKAQVAAEKA 1612

Qy     380 VEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATD-- 437
           ::|   : : | : |  :         | :  | : : |:   ||:      : |   :
Db    1613 IKQADEDIQGTQNLLTSIESETAASEETLFNASQRISELERNVEELKRKAAQNSGEAEYI 1672

Qy     438 EQAV-----STEEVAEMVDG-VDDRAGE----IAAALDDIADA 470
           |: |     | |:| : :|| :|::  :    ||   :: |||
Db    1673 EKVVYTVKQSAEDVKKTLDGELDEKYKKVENLIAKKTEESADA 1715


Search completed: January  2, 2003, 12:40:02
Job time : 21 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:36:16 ; Search time 96 Seconds
(without alignments)
678.745 Million cell updates/sec

Title: US-09-455-978B-2
Perfect score: 2394
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVRTVEEVRETVGKLS 489

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2394	100.0	489	22	AAE04641	Halobacterium sali
2	412.5	17.2	432	22	AAE04642	Bacillus subtilis
3	387.5	16.2	739	22	AAB96493	Putative sensory t
4	380	15.9	435	22	AAB96483	Putative sensory t
5	375	15.7	501	22	AAB96709	Putative sensory t
6	359.5	15.0	891	21	AAB26596	Synechocystis sp p
7	356.5	14.9	374	22	AAB96382	Putative P. abyssi
8	339.5	14.2	601	23	ABB49909	Listeria monocytog
9	298	12.4	511	21	AAB14148	Bordetella pertuss
10	297.5	12.4	1137	22	ABG24223	Novel human diagno
11	289	12.1	565	19	AAW98798	H. pylori GHPO 128
12	289	12.1	565	19	AAW71554	Helicobacter polyp
13	289	12.1	565	22	AAB46327	H. pylori HPS024 p
14	284	11.9	630	19	AAW98797	H. pylori GHPO 127
15	284	11.9	630	19	AAW71553	Helicobacter polyp
16	284	11.9	675	22	AAB46345	H. pylori HPS149 p
17	280	11.7	675	19	AAY10961	H. pylori ORF 09ae
18	276	11.5	293	18	AAW20287	H. pylori inner me
19	273.5	11.4	664	18	AAW20769	H. pylori inner me
20	269	11.2	723	20	AAY19897	B. burgdorferi ant
21	269	11.2	753	20	AAY19896	B. burgdorferi ant
22	266	11.1	606	20	AAY19800	B. burgdorferi ant
23	266	11.1	633	20	AAY19799	B. burgdorferi ant
24	257	10.7	431	21	AAB52596	Helicobacter pylor
25	254	10.6	433	17	AAW05196	Helicobacter pylor
26	253	10.6	300	23	ABB49349	Listeria monocytog
27	252.5	10.5	438	18	AAW20916	H. pylori inner me
28	249.5	10.4	701	22	ABG25139	Novel human diagno
29	242.5	10.1	191	22	ABG18218	Novel human diagno
30	221	9.2	566	22	ABG18219	Novel human diagno
31	214.5	9.0	110	23	ABP09757	Human ORFX protein
32	207.5	8.7	524	20	AAY20011	B. burgdorferi ant
33	207.5	8.7	553	20	AAY20010	B. burgdorferi ant
34	203	8.5	39	22	AAE04677	Haem protein relat
35	202.5	8.5	955	15	AAR57365	K39 polypeptide of
36	202.5	8.5	955	17	AAW03691	Leishmania chagasi
37	199	8.3	258	22	AAB96439	Putative sensory t
38	198	8.3	98	21	AAB52611	Helicobacter pylor
39	196.5	8.2	1857	23	AAU84350	Protein MYH11 diff
40	194.5	8.1	1851	22	ABG01723	Novel human diagno
41	194.5	8.1	1960	22	AAM78854	Human protein SEQ
42	194.5	8.1	2143	22	ABG01716	Novel human diagno
43	194	8.1	1786	18	AAW24790	P. falciparum live
44	194	8.1	1787	23	AAU96699	Plasmodium falcipo
45	194	8.1	3854	22	ABB67104	Drosophila melanog

ALIGNMENTS

RESULT 1
AAE04641
ID AAE04641 standard; Protein; 489 AA.
XX
AC AAE04641;
XX
DT 04-SEP-2001 (first entry)
XX
DE Halobacterium salinarium HemAT-Hs protein.
XX
KW Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;
KW oxygen storage; artificial photosynthesis; signalling function;
KW alpha-haemoglobin; myoglobin; therapy.
XX
OS Halobacterium salinarium.
XX
PN WO200140475-A2.
XX
PD 07-JUN-2001.
XX
PF 05-DEC-2000; 2000WO-US33048.
XX
PR 06-DEC-1999; 99US-0455978.
XX
PA (UYHA-) UNIV HAWAII.
XX
PI Alam M, Larsen R;
XX
DR WPI; 2001-374832/39.
DR N-PSDB; AAD08991.
XX
PT Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs
PT which reversibly binds oxygen with low affinity, useful for controlled

```
PT   storage of oxygen and for sensing gaseous ligands such as oxygen  -
XX
PS   Claim 6; Page 10; 94pp; English.
XX
CC   The present invention relates to isolated archeal and bacterial haem
CC   binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen
CC   with low affinity. Haem binding protein is useful for controlled storage
CC   of oxygen by allowing haem binding protein to bind and store oxygen, and
CC   triggering the release of oxygen from haem binding protein by activating
CC   the signalling domain. Haem binding protein is useful for sensing gaseous
CC   ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem
CC   binding protein is useful for treating a patient suffering from low
CC   blood levels by administering and regulating the oxygen binding of the
CC   haem-binding protein by modifying the signalling domain. Haem binding
CC   protein is useful for haem-based catalysis, for artificial
CC   photosynthesis and for identifying potential signalling functions of
CC   mutated alpha-haemoglobin and myoglobin causing several diseases.
CC   The present sequence is Halobacterium salinarium HemAT-Hs protein which
CC   is salt tolerant.
XX
SQ   Sequence   489 AA;

  Query Match           100.0%;  Score 2394;  DB 22;  Length 489;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-168;
  Matches  489;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSNDNDTLVTADVRNGIDGHALADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQPLFEAT 60

Qy      61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIG 120

Qy     121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 KIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTF 180

Qy     181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRART 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRART 240

Qy     241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 DDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDID 300

Qy     301 EATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 EATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVAD 360

Qy     361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 EVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITD 420

Qy     421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEE 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEE 480

Qy     481 VRETVGKLS 489
          |||||||||
Db     481 VRETVGKLS 489


RESULT 2
AAE04642
ID   AAE04642 standard; Protein; 432 AA.
XX
AC   AAE04642;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Bacillus subtilis HemAT-Bs protein.
XX
KW   Haem binding protein; HemAT-Hs; HemAT-Bs; gaseous ligand sensor;
KW   oxygen storage; artificial photosynthesis; signalling function;
KW   alpha-haemoglobin; myoglobin; therapy.
XX
OS   Bacillus subtilis.
XX
PN   WO200140475-A2.
XX
PD   07-JUN-2001.
XX
PF   05-DEC-2000; 2000WO-US33048.
XX
PR   06-DEC-1999;   99US-0455978.
XX
PA   (UYHA-) UNIV HAWAII.
XX
PI   Alam M,  Larsen R;
XX
DR   WPI; 2001-374832/39.
DR   N-PSDB; AAD08992.
XX
PT   Novel isolated bacterial heme binding protein, HemAT-Hs or HemAT-Bs
PT   which reversibly binds oxygen with low affinity, useful for controlled
PT   storage of oxygen and for sensing gaseous ligands such as oxygen  -
XX
PS   Claim 8; Page 11; 94pp; English.
XX
CC   The present invention relates to isolated archeal and bacterial haem
CC   binding protein, HemAT-Hs or HemAT-Bs which reversibly binds oxygen
CC   with low affinity. Haem binding protein is useful for controlled storage
CC   of oxygen by allowing haem binding protein to bind and store oxygen, and
CC   triggering the release of oxygen from haem binding protein by activating
CC   the signalling domain. Haem binding protein is useful for sensing gaseous
CC   ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem
CC   binding protein is useful for treating a patient suffering from low
CC   blood levels by administering and regulating the oxygen binding of the
CC   haem-binding protein by modifying the signalling domain. Haem binding
CC   protein is useful for haem-based catalysis, for artificial
CC   photosynthesis and for identifying potential signalling functions of
CC   mutated alpha-haemoglobin and myoglobin causing several diseases.
CC   The present sequence is Bacillus subtilis HemAT-Bs protein.
XX
SQ   Sequence   432 AA;

  Query Match           17.2%;  Score 412.5;  DB 22;  Length 432;
  Best Local Similarity  26.2%;  Pred. No. 2.4e-22;
  Matches  114;  Conservative   85;  Mismatches  187;  Indels   49;  Gaps    8;

Qy      31 AEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTKTVE 90
           |::  :|   : |  :  |   ||| :     :|  || :|:      |: |   :|:
Db      32 ADVKKQLKMVRLGDAELYVLEQLQPLIQENIVNIVDAFYKNLDHESSLMDII-NDHSSVD 90

Qy      91 QLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALAD 150
           :||:|   ::  :  |  | |:  :| ||  ||  :|| |  |:||:      ::|
Db      91 RLKQTLKRHIQEMFAGVIDDEFIEKRNRIASIHLRIGLLPKWYMGAFQELLLSMIDIY-- 148

Qy     151 DVVADRGEEAAAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANA 210
                 | :     ||:       |:|  :||: :: :   | |   || : :: | :
Db     149 -------EASITNQQELLKAIKATTKILNLEQQLVLEAFQSEYNQ-TRDEQEEKKNLLH- 199

Qy     211 VATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADD 270
                    :: ||  :|     : : |   |  : | |  ||  | :    :|| ::
Db     200 ---------QKIQETSGSIA----NLFSETSRSVQELVDKSEGISQASKAGTVTSSTVEE 246

Qy     271 VRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDI 330
           :      |   || :         |  || :   :   : :| | |  :| : |::  |
Db     247 -KSIGGKKELEVQQKQ---------MNKIDTSLVQIEKEMVKLDEIAQQIEKIFGIVTGI 296

Qy     331 AEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEET 390
           |||||:|:||||||:|||||  |:||||||:||: |:|::::  : | |||     :
Db     297 AEQTNLLSLNASIESARAGEHGKGFAVVANEVRKLSEDTKKTVSTVSELVNNTNTQINIV 356

Qy     391 VDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDA-------------ASGMQEVSTAT 436
```

```
Db    357 SKHIKDVNELVSESKEKMTQINRLFDEIVHSMKISKEQSGKIDVDLQAFLGGLQEVSRAV 416

Qy    437 DEQAVSTEEVAEMVD 451
Db    417 SHVAASVDSLVILTE 431
```

RESULT 3
AAB96493
ID AAB96493 standard; Protein; 739 AA.
XX
AC AAB96493;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative sensory transduction histidine kinase and response regulator #3.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-0005034.
XX
PR 21-APR-1999; 99FR-0005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
DR WPI; 2001-126236/14.
XX
PT New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
XX
PS Claim 7; Pages 1203-1205; 1657pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO200065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.
XX
SQ Sequence 739 AA;

Query Match 16.2%; Score 387.5; DB 22; Length 739;
Best Local Similarity 28.6%; Pred. No. 3.4e-20;
Matches 126; Conservative 83; Mismatches 150; Indels 81; Gaps 15;

```
Qy     83 ANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQ-----RARIGKI--------HDVLGLG 129
Db    331 SNTLAPLEKLK--YAAQALAEGRLKQVSEYLKQIRYLERDEIGALIQAFEAVSKDLVGT- 387

Qy    130 PDVYLGAYTRYYTGLLDALADD------VVADRGEEAAAAVDELVARFLPMLKLLTFDQQ 183
Db    388 ----LNAISK----KLERLAEGDLSNGLTVEVRGE---------------LRDIIED-- 421

Qy    184 IAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQ 243
Db    422 ------IRSVTETFRESIGSLVEMAN----DLEKRANALAQVSKDVTEAINQVN-EAIQQ 470

Qy    244 VDRMADVSRE-ISSVSASVEEVASTADDVRRTSE-------DAEALAQQGEAAADDALAT 295
```

```
Db    471 VSIEAQRQQETINEITDGMRLVAQTSEESVRAMEEFSGAVTEVVSIANEGSQKGDEALKR 530

Qy    296 MTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGF 355
Db    531 IEDIQHMMSRIEETVSKVAEMSRNIEEITNVITSIAEQTNLLALNAAIEAARAGEAGRGF 590

Qy    356 AVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETL 415
Db    591 AVVAQEIRKLAEESKQAADNIKSIIDKITDEIKEAVEATK-------EGVSVIGESSETL 643

Qy    416 QEITDAVEDAASGMQEVS---TATDEQAVSTEEVAE----MVDGVDDRAGEIAAALDDIA 468
Db    644 RDTIGYLANIATLLQETSERMTTVKEQIVRTQEEVDKALRALENLAASAEETTASAEEVS 703

Qy    469 DATDQQVRTVEEVRETVGKL 488
Db    704 SAIEQQTAAIEELRRAAQEL 723
```

RESULT 4
AAB96483
ID AAB96483 standard; Protein; 435 AA.
XX
AC AAB96483;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative sensory transduction histidine kinase and response regulator #2.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-0005034.
XX
PR 21-APR-1999; 99FR-0005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
DR WPI; 2001-126236/14.
XX
PT New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
XX
PS Claim 7; Pages 1189-1190; 1657pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO200065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.
XX
SQ Sequence 435 AA;

Query Match 15.9%; Score 380; DB 22; Length 435;
Best Local Similarity 27.5%; Pred. No. 6.1e-20;
Matches 109; Conservative 91; Mismatches 129; Indels 68; Gaps 12;

```
Qy    107 EYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDE 166
```

```
Db     63 EYEKQLQREIDEIVKILDRIAQG-DL-------------SVEDKEIGGMGAKVRQGIED 107

Qy    167 LVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAV-ATHVEAPLSSLEAT 225
Db    108 L-RKSISQLVLNIKNAAIDVRNHTKVMKENIEQVADSVQQVAEAINQVSIEA-----QRE 161

Qy    226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
Db    162 QENISKMTDTMR--------YIHDISKE--TVSTMEEFEAS----MREMAQLAKEGGEKG 207

Qy    286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
Db    208 RQAADQ-IEEISRMMEKIEETVRGVAEMGK---SIENITNVISSISEQTNLLALNAAIEA 263

Qy    346 ARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGV 405
Db    264 ARAGEAGKGFAVVADEIRKLAEESKKAAEDIRELIKQ-------------IGDKIGESV 309

Qy    406 ERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMV-------------D 451
Db    310 EVTQQGAEVVKTSTEVIKESVSYLTQVAEMMEEMEVKASELREKVIQEGEKIEEGLRFLE 369

Qy    452 GVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKL 488
Db    370 NLAASAEETTAAAEEVSAAAEEQSSALQELRESVKEL 406


RESULT 5
AAB96709
ID   AAB96709 standard; Protein; 501 AA.
XX
AC   AAB96709;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Putative sensory transduction histidine kinase & response regulator #4.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS   Pyrococcus abyssi.
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 7; Pages 1469-1471; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC   a hyperthermophilic archaeon, which is isolated from deep-sea
CC   hydrothermal vents. The present sequence is one such P. abyssi protein.
CC   The proteins of the present invention have various potential industrial
CC   uses, since the proteins are stable at very high temperatures, some up to
CC   110 degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence   501 AA;

  Query Match             15.7%;  Score 375;  DB 22;  Length 501;
  Best Local Similarity   24.4%;  Pred. No. 1.7e-19;
  Matches  116;  Conservative  111;  Mismatches  175;  Indels   74;  Gaps   12;

Qy     65 VTDFYDHLE----------SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAA 114
Db     30 ISDLYDNLEKTMKVVQTGGSVEAIRETLENAK---SQLKET-----LWLSIGVMSI---- 77

Qy    115 QRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPM 174
Db     78 ----VAVISGALGM----RLMNSTMRPINEMAKIAESIAEGKLSRAREMVSKIQYREDDE 129

Qy    175 L-KLLTFDQQIAMDTY--IDSYAQRLHDEIDSRQELANAVATHV----EAPLSSLEATSQ 227
Db    130 IGKLIEGFRAISQDVLQTLEIITERM--EKISEGDVSDDLKVHAKGDFESILNSMRKTIS 187

Qy    228 DVAERTDTMR-------ARTDDQVDRMADVSREISSVSASVEEVA--------------- 265
Db    188 NLRELMKTVRDLALTLESRANDLTRISSEISEAINQVAEAIQQVSVEAQRQQENITEIME 247

Qy    266 ---STADDVRRTSEDAE----------ALAQQGEAAADDALATMTDIDEATDGVTAGVEQ 312
Db    248 GMNITADVTQRTVDAMEEFSGVVNEVLSIAREGKDKGEKAISQVEDIQDAMKVIRQAVQE 307

Qy    313 LGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQ 372
Db    308 VAEMSKNVGDIINAIADIAEQTNLLALNAAIEAARAGELGRGFAVVAQEVRNLAEESKEA 367

Qy    373 STRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEV 432
Db    368 AEKIRGILNEIQEKVEKAVEETEKGVKVVDDSVDFLKETVGYLMNIGELLDDVESKLQDI 427

Qy    433 STATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGKL 488
Db    428 KNELANTQEHVENAKKALENLAASAQETTASAEEVSASAQEQASSMEEVKRNIIEL 483


RESULT 6
AAB26596
ID   AAB26596 standard; protein; 891 AA.
XX
AC   AAB26596;
XX
DT   01-FEB-2001  (first entry)
XX
DE   Synechocystis sp phytochrome-related gene Cph5.
XX
KW   Phytochrome; fluorescent apophytochrome; bilin; Cph.
XX
OS   Synechocystis sp.
XX
PN   WO200056355-A1.
XX
PD   28-SEP-2000.
XX
PF   14-MAR-2000; 2000WO-US06607.
XX
PR   19-MAR-1999;   99US-0272809.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Lagarias JC;
XX
DR   WPI; 2000-602195/57.
XX
PT   Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
PT   Synechocystis species) and phycoerythrobilin conjugate, useful as
PT   fluorescent markers for biological research -
XX
PS   Example 1; Page 48; 52pp; English.
```

```
XX
CC   The present invention is related to fluorescent apophytochrome-bilin
CC   conjugates, known as phytofluors. An apoprotein known as Cph2 from
CC   Synechocystis species is used as the apophytochrome and the bilin is
CC   preferably phycoerythrobilin. The phytofluors are useful as fluorescent
CC   markers for biological research. The phytofluors have a long wavelength
CC   absorption maxima, a high molar absorption coefficient and the
CC   recombinant apoproteins can spontaneously assemble with a variety of
CC   bilin chromophore precursors. The present sequence is a phytochrome
CC   related protein from Synechocystis sp.
XX
SQ   Sequence   891 AA;

  Query Match             15.0%;  Score 359.5;  DB 21;  Length 891;
  Best Local Similarity   27.1%;  Pred. No. 5.1e-18;
  Matches  136;  Conservative   86;  Mismatches  197;  Indels   83;  Gaps   19;

Qy      23 ADRIGLDEAEIAWRLSFTGIDDDTMAALAAEQP--LFEATADALVTDFYDHLESY----- 75
           |||:      | :|   |      : ::|   | |    ||    | | :| |
Db     418 ADRV------IVYRFDATWAGTVIVESVAEGYPKALGATIADPCFADSY--VEKYRSGRI 469

Qy      76 ERTQDLFANSTKT---VEQLK--ETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGL-- 128
           : |:|:: |:  |   : |||  | :|  :                | |    ::|||
Db     470 QATRDIY-NAGLTPCHIGQLKPFEVKANLV---------------APINYKGNLLGLLI 512

Qy     129 -----GP------DVYLGAYTRYYTGLL----DALADDVVADRGEEAAAAVDELVARFLP 173
                ||      :: |       ||     | ||   :|:   |     :::  | |
Db     513 AHQCSGPRDWHQNEIDLFGQLTVQVGLALERSDLLAQQKIAE--VEQRQMREKMQKRALE 570

Qy     174 MLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERT 233
           :  |:  |      |  | ::     ||| :  :  ||    :   :: ::  :    | |
Db     571 L--LMEVDPVSRGDLTIRAHV--TEDEIGTIADSYNATIESLRRIVTQVQTAASQFTETT 626

Qy     234 DT-------MRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGE 286
           ||       :  : : |   :|:    : ::: |:: ||  |       : |     |||
Db     627 DTNEVAVRQLAQQANRQALDVAEALERLQAMNKSIQAVAENAAQAESAVQRATQTVDQGE 686

Qy     287 AAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAA 346
           | :  :  :  | |        |::||| :  :  |  :|   |:|||:|||||:||||
Db     687 DAMNRTVDGIVAIRETVAATAKQVKRLGESSQKISKVVNLIGSFADQTNLLALNAAIEAA 746

Qy     347 RAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVE 406
           ||| | |||||||||::|| :| | :  : :||  :|||| | |: ::   :::  | :
Db     747 HAGEEGRGFAVVADEVRSLARQSAEATAEIAQLVATIQAETNEVVNAMEAGTEQVVVGTK 806

Qy     407 RVEEAMETLQEITDAVEDAASGMQE-VSTATDEQAVSTEE-------VAEMVD------- 451
           |||   :| :|| ||    ||: | :::|  ||: ::|        ||:: |
Db     807 LVEETRRSLNQIT-AVSAQISGLVEAITSAAIEQSQTSESVTQTMALVAQIADKNSSEAS 865

Qy     452 GVDDRAGEIAAALDDIADATDQ 473
           ||     |: |    : :|  |
Db     866 GVSATFKELLAVAQSLQEAVKQ 887


RESULT 7
AAB96382
ID   AAB96382 standard; Protein; 374 AA.
XX
AC   AAB96382;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Putative P. abyssi sensory transduction histidine kinase.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS   Pyrococcus abyssi.
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 7; Pages 1066-1067; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC   a hyperthermophilic archaeon, which is isolated from deep-sea
CC   hydrothermal vents. The present sequence is one such P. abyssi protein.
CC   The proteins of the present invention have various potential industrial
CC   uses, since the proteins are stable at very high temperatures, some up to
CC   110 degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence   374 AA;

  Query Match             14.9%;  Score 356.5;  DB 22;  Length 374;
  Best Local Similarity   26.6%;  Pred. No. 2.7e-18;
  Matches  105;  Conservative   85;  Mismatches  145;  Indels   59;  Gaps   10;

Qy     104 GRGEYD-TEYAAQRARIGKIHDVL-GLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAA 161
           |: :|: :     :| |  : ||| |  | : ||                  ::   ||
Db      19 GKKDYENSSNVTSQAIINYLRDVLDGKEPQIPLG----------------LSKEDEE-- 59

Qy     162 AAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSS 221
                     :|| : |  :       |:     : |:|:: :|    |  ::|  :
Db      60 -----------VLKKVAFRLKRGQGKKIN-----VKDKIENLKE----VIENLEEKIG- 97

Qy     222 LEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEAL 281
           |    |: |  : : :| :::  ::|:|:  | ::|| :||:   |  :   : :: ::
Db      98 -EVKLGDLGE-VNELVSRLNNENVKIAEVNDYIQTLSAGIEEMNVQAQQLSDFALESASM 155

Qy     282 AQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNA 341
           |::|   :|:    :: | | :  ::  |  | |: :  :  :  ||  || |||:|||||
Db     156 AEKGRQISDNVALKVSRISETSREMSDAVRILAEYSKKINDIVYVISSIASQTNLLALNA 215

Qy     342 SIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQRI 401
           |||||||||||| ||||||: |: ||: |:: :  :  |:|:||       : :: | | |
Db     216 SIEAARAGEAGRGFAVVAENVRELADRSKKSAREIRNLIEEMQ-------ENINRVIQAI 268

Qy     402 GEGVERVEEAMETLQEITDAVED-------AASGMQEVSTATDEQAVSTEEVAEMVDGVD 454
           | |   ||  | :| :  | :|         |: ::|:|   ||||  : : : :| :
Db     269 QENVRVTEEVKEAIQNLIAAFDDIARRANETANMVKELSEGIDEQANSVQMLVDRIDSIS 328

Qy     455 DRAGEIAAALDDIADATDQQVRTVEEVRETVGKL 488
                :       : |      :  : ||:| : ||
Db     329 KDVSDNLNFATQLTDTISGSLEKLNEVKEEITKL 362


RESULT 8
ABB49909
ID   ABB49909 standard; Protein; 601 AA.
XX
AC   ABB49909;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Listeria monocytogenes protein #2613.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
```

```
KW   vitamin B12; bacterial infection; disease.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI   Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI   Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI   Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI   Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI   Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI   Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI   Rose M, Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides -
XX
PS   Claim 6; SEQ ID No 2614; 192pp; French.
XX
CC   The present invention relates to the genome sequence of Listeria
CC   monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC   it are useful for selecting probes and primers for detecting genes in L.
CC   monocytogenes and related organisms, and for studying genetic
CC   polymorphisms and other genomes. The present sequence is a protein
CC   encoded by the genome sequence of the present invention. Proteins
CC   expressed from the genome sequence are useful for raising specific
CC   antibodies, identification of L. monocytogenes and related organisms, and
CC   for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC   B12. The genome sequence and proteins encoded by it are also useful for
CC   selecting compounds that regulate gene expression and cell replication
CC   and modulate L. monocytogenes-related diseases. In addition, the genome
CC   sequence and proteins encoded by it are useful in pharmaceutical and
CC   vaccines compositions for the treatment or prevention of infections by L.
CC   monocytogenes and related organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   601 AA;

  Query Match           14.2%;  Score 339.5;  DB 23;  Length 601;
  Best Local Similarity  24.0%;  Pred. No. 9.1e-17;
  Matches  125;  Conservative   93;  Mismatches  224;  Indels   79;  Gaps   11;

Qy      31 AEIAWRLSFTGIDDDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTKTVE 90
Db      57 AKIQTNMAQINIDILTMFDTINGKSTLIKDIDNLYAENDQAIHNFKKA-NLTAEDKKQLA 115

Qy      91 QLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYY--------- 141
Db     116 YFEEKLADMKASASSVISDTSSALDDAEL-------------LGAQNRYYQNVKTKFDD 161

Qy     142 -TGLLDALAD---------DVVADRGEE-----AAAAVDELVARFL-------PMLK-- 176
Db     162 ATKQLNVLNDMNYKEVENSSQAISDFGVKISLIFTAVIIAVLISLFIFNAYITRVILKGI 221

Qy     177 --LLTFDQQIAMD--TYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAER 232
Db     222 RHLQTAVHKVASGDLSYRSTYNGR--DELGDITNDLNEMSENLRLMIEDVKKASTDVKSS 279

Qy     233 TDT----------------------MRARTDDQVDRMADVSREISSVSASVEEVASTADDV 271
```

```
Db     280 SDNVIISSEIISAMTTEMDIEMKMMGEQIQTQIGSMKESTDAMDQMTGGVQNVAEYALKV 339

Qy     272 RRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIA 331
Db     340 SDLTKDSAEKTNDGIAVINNLVSQMDRISGVMRSSTDVVSQLVNRVGEVEKALDTVTNIA 399

Qy     332 EQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETV 391
Db     400 DQTNLLALNAAIESARAGEHGRGFAVVAEEVRKLAEQSRLAVVDINTVLKKIQTESKTTI 459

Qy     392 DQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVD 451
Db     460 EVMNTGLSESEAGQKIISETEATFTDLLNRVNDISAQMQNVSQETEEMAAGIEEVNTSIS 519

Qy     452 GV---DDRAGEIAAALDDIADATDQQVRTVEEVRETVGKLS 489
Db     520 DVTEISNQIGEKSTAALEFAEVNKMKVDELVVISEEMQKIS 560
```

```
RESULT 9
AAB14148
ID   AAB14148 standard; Protein; 511 AA.
XX
AC   AAB14148;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Bordetella pertussis protein # 2.
XX
KW   Type III secretion system; virulence factor; pathogenicity island.
XX
OS   Bordetella pertussis.
XX
PN   WO200037493-A2.
XX
PD   29-JUN-2000.
XX
PF   21-DEC-1999;   99WO-EP10297.
XX
PR   21-DEC-1998;   98GB-0028217.
XX
PA   (ULBR ) UNIV LIBRE BRUXELLES.
XX
PI   Bollen A,  Fauconnier A,  Godfroid E;
XX
DR   WPI; 2000-452178/39.
DR   N-PSDB; AAA64886, AAA64890.
XX
PT   Novel polypeptides derived from Bordetella pertussis, useful for
PT   treating and diagnosing Bordetella infection -
XX
PS   Disclosure; Pages 156-157; 165pp; English.
XX
CC   Bordetella pertussis possesses a type III secretion system. Type III
CC   secretion systems allow bacteria to target virulence factors directly at
CC   host cells. The present sequence is a protein of B. pertussis. No name
CC   has been specified for the present sequence. The coding sequence of the
CC   present sequence is homologous to housekeeping genes of other species,
CC   and so the present protein may have a role in cellular housekeeping. A
CC   pathogenicity island is a compact, distinct genetic unit carrying
CC   virulence genes. The coding sequence of the present sequence is located
CC   within a pathogenicity island (see AAA64890) which also carries, a
CC   number of genes encoding proteins involved in the type III secretion
CC   system of B. pertussis. See AAA64849-A64884 and AAB14111-B14146 for
CC   details of the coding sequences and proteins identified in the
CC   pathogenicity island, of the present invention.
XX
SQ   Sequence   511 AA;

  Query Match           12.4%;  Score 298;  DB 21;  Length 511;
  Best Local Similarity  24.0%;  Pred. No. 8.4e-14;
  Matches  118;  Conservative   87;  Mismatches  184;  Indels  102;  Gaps   15;
```

```
Qy      33 IAWR---LSFTGIDDDTMAALAAEQPLFEATADA---LVTDFYDHLE---------- 73
           : ||    |   |:     |:    |||||:  |     | |  :::
Db      28 LTWRNAGRSAAEIEGLNQVAVNQVDPLFEASGAAQRQAATQFQRYVDVPKEPAAAELAAT 87

Qy      74 ------SYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLG 127
                :|:   |  | :      :   |||  |          |:| ||
Db      88 LQTRWRAYQSVLDELAAA------VDAGQAEPALAAMHRAQQAEHAFQRD---------- 131

Qy     128 LGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVD-ELVARF----LPMLKLL---- 178
                          ::|    | |   |  : | |  :|||:    |  | :|
Db     132 ----------------MEAFLARVQAHSDEVRSGAEDTHVVARWSAIALTTLGVLLTLA 174

Qy     179 --TFDQQIAMDTYIDS--YAQRLHD-EIDSRQELANAVATHVEAPLSSLEATSQDVAERT 233
             | ::  :   :::  :  |: | :: :| |: :  |  : |  ::|:   : :
Db     175 GWLFVRRAVLRPLLEAGHHFDRIADGDLTARIEVRS--ANEIGALFAALKRMQEGLTRTI 232

Qy     234 DTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG--EAAADD 291
             ||   |:     |::|   :::|:  || |:  ::   | |:     :|    ||  :
Db     233 AVMRRGVDEINVGAAEISAGNANLSSRTEEQAAALEETAATMEELATTVKQNADNAAQAN 292

Qy     292 ALA------------TMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLAL 339
           ||             ::  : :   |::|   |: :        :  ||| || |||:|||
Db     293 QLAAVSMQVAQRGGESVAQVVQTMHGISASSRQIAD-------IVTVIDGIAFQTNILAL 345

Qy     340 NASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAETEETVDQLDEVNQ 399
           ||::|||||| |:|||||| ||::||: : : :  :: |:|   |       |:
Db     346 NAAVEAARAGEQGKGFAVVAGEVRSLAQRAAQAAKEIKALIESSVATVRAGSQQVASAGG 405

Qy     400 RIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTA-------TDEQAVSTEEVAEMVDG 452
             : | |  |:   : : ||: |    |||: :|| |       | : |   |: |
Db     406 TMDEVVASVQRVADIMGEISAASAQQASGIDQVSLAISQMDETTQQNAALVEQAAAAATA 465

Qy     453 VDDRAGEIAAA 463
           ::::|  :|||
Db     466 MEEQARHLAAA 476


RESULT 10
ABG24223
ID   ABG24223 standard; Protein; 1137 AA.
XX
AC   ABG24223;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #24214.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS88410.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 54582; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1137 AA;

  Query Match             12.4%;  Score 297.5;  DB 22;  Length 1137;
  Best Local Similarity   29.0%;  Pred. No. 2.6e-13;
  Matches   85;  Conservative   61;  Mismatches  110;  Indels   37;  Gaps    7;

Qy     204 RQELANAVATHVEAPLSSLEATSQDVAERTDTMR-------ARTDDQVDRMADVSREISS 256
           | |:  |:   ::    :|  |  || : :  |         |  :|   |    :  ::
Db      30 RNEI-TAIFASLKTMQQALRGTVSDVRKGSQEMHIGIAEIVAGNNDLSSRTEQQAASLAQ 88

Qy     257 VSASVEEVAST----ADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQ 312
            :||:|:: :|    ||: |: || |:  |   :|     ::|||           :::
Db      89 TAASMEQLTATVGQNADNARQASELAKNAATTAQAGGVQ-VSTMTHT----------MQE 137

Qy     313 LGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQ 372
           :   :  :  :  ||| || |||:||||::||||||| | |||||| ||: ||   | :
Db     138 IATSSQKIGDIISVIDGIAFQTNILALNAAVEAARAGEQGRGFAVVAGEVRNLASRSAQA 197

Qy     373 STRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEV 432
           :  :: |:|                : || |: :| : |  |  |: :|  :|      | |:
Db     198 AKEIKGLIE------------ESVN-RVQQGSKLVNNAAATMIDIVSSVTRVNDIMGEI 243

Qy     433 STATDEQAVSTEEVAEMVDGVDDRAGEIAAALDDIADATDQQVRTVEEVRETV 485
           ::|::||    |:||: |  :|    : |: ::: | ||:|     : :   |
Db     244 ASASEEQQRGIEQVAQAVSQMDQVTQQNASLVEEAAVATEQLANQADHLSSRV 296


RESULT 11
AAW98798
ID   AAW98798 standard; Protein; 565 AA.
XX
AC   AAW98798;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 1282 protein.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
XX
OS   Helicobacter pylori.
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
```

```
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   N-PSDB; AAX14517.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 8; Page 1743-1745; 2054pp; English.
XX
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence   565 AA;

  Query Match           12.1%;  Score 289;  DB 19;  Length 565;
  Best Local Similarity   27.2%;  Pred. No. 4.4e-13;
  Matches   90;  Conservative   69;  Mismatches  122;  Indels   50;  Gaps    9;

Qy     164 VDELVARFLPMLKLLTF-----DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVE-- 216
           :||||      ||:  |     | :  :|         : |  |   ::   :   ||
Db     239 IDELV------LKINAFSRGDKDLRAKID---------VGDRNDEISQVGRGINLFVENA 283

Qy     217 -------APLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVAS--- 266
                 :|:|  || |   :       | :    | | |  ::||     ::||
Db     284 RLIMEEIKGISTLNKTSMDKLVQI------TQETQKSMKDSSTTLNSVKNKATDIASMMN 337

Qy     267 ----TADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVES 322
               :  :|:  :  : : | :: : |  |  : :|:     : ::: |||   | ||:|
Db     338 ASIEQSQGLRKRLIETQGLVKESKDAIGDLFSQITESAHTEEELSSKVEQLSRNADDVKS 397

Qy     323 VTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQ 382
           :  :|:|||:||:|||||:|||||| | |||||||||: ||  :::   :   :
Db     398 ILDIINDIADQTNLLALNAAIEAARAGEHGRGFAVVADEVRNLAGRTQKSLAEINSTIMV 457

Qy     383 MQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAV-EDAASGMQEVSTATDEQAV 441
           :  |      |::  :|:    :||: :  :::||  : :  : :| : : : : |: |
Db     458 IVQEINAVSSQMNLNSQK----MERLSDMSKSVQETYEKMSSNLSSVVSDSNQSMDDYAK 513

Qy     442 STEEVAEMVDGVDDRAGEIAAALDDIADATD 472
           |  ::  |   | || |     |   :||::|
Db     514 SGHQIEVM---VSDFAEVEKVASKTLADSSD 541


RESULT 12
AAW71554
ID   AAW71554 standard; Protein; 565 AA.
XX
AC   AAW71554;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Helicobacter polypeptide GHPO 1282.
XX
KW   GHPO 1282; infection; therapy; diagnosis; vaccine; gastritis;
KW   ulcer.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 123
FT                   /note= "encoded by GNC"
XX
PN   WO9821225-A1.
XX
PD   22-MAY-1998.
XX
PF   14-NOV-1997;   97WO-US21353.
XX
PR   29-JUL-1997;   97US-0902615.
PR   14-NOV-1996;   96US-0749051.
PR   01-APR-1997;   97US-0831309.
PR   01-APR-1997;   97US-0833457.
PR   01-APR-1997;   97US-0834705.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Haas R,  Kleanthous H,  Meyer T,  Miller C;
PI   Odenbreit S,  Tomb J;
XX
DR   WPI; 1998-297855/26.
DR   N-PSDB; AAV52089.
XX
PT   Helicobacter polynucleotide and polypeptide sequences - useful to
PT   treat or prevent gastrointestinal infection
XX
PS   Claim 1; Page 313-314; 362pp; English.
XX
CC   This claimed Helicobacter pylori polypeptide, designated GHPO 1282,
CC   can be used in vaccination methods for preventing or treating
CC   Helicobacter infection.  85 Helicobacter polypeptides (see
CC   AAW71474-W71558) are claimed, as well as isolated polynucleotides
CC   (see AAV52009-93) that encode them.  The invention also provides:
CC   methods for producing these Helicobacter polypeptides in
CC   recombinant host systems, and related expression cassettes, vectors
CC   and transformed or transfected host cells; live vaccine vectors
CC   that contain the polynucleotides of the invention and which can be
CC   used to prevent or treat Helicobacter infection; therapeutic and/or
CC   prophylactic methods involving administration of polynucleotide
CC   molecules, polypeptides or monospecific antibodies; methods for
CC   detecting the presence of Helicobacter in samples using e.g.
CC   the polypeptides or monospecific antibodies; and methods for
CC   purifying the polypeptides by antibody-based affinity
CC   chromatography.
XX
SQ   Sequence   565 AA;

  Query Match           12.1%;  Score 289;  DB 19;  Length 565;
  Best Local Similarity   27.2%;  Pred. No. 4.4e-13;
  Matches   90;  Conservative   69;  Mismatches  122;  Indels   50;  Gaps    9;

Qy     164 VDELVARFLPMLKLLTF-----DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVE-- 216
           :||||      ||:  |     | :  :|         : |  |   ::   :   ||
Db     239 IDELV------LKINAFSRGDKDLRAKID---------VGDRNDEISQVGRGINLFVENA 283

Qy     217 -------APLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVAS--- 266
                 :|:|  || |   :       | :    | | |  ::||     ::||
Db     284 RLIMEEIKGISTLNKTSMDKLVQI------TQETQKSMKDSSTTLNSVKNKATDIASMMN 337

Qy     267 ----TADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVES 322
               :  :|:  :  : : | :: : |  |  : :|:     : ::: |||   | ||:|
Db     338 ASIEQSQGLRKRLIETQGLVKESKDAIGDLFSQITESAHTEEELSSKVEQLSRNADDVKS 397

Qy     323 VTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQ 382
           :  :|:|||:||:|||||:|||||| | |||||||||: ||  :::   :   :
Db     398 ILDIINDIADQTNLLALNAAIEAARAGEHGRGFAVVADEVRNLAGRTQKSLAEINSTIMV 457

Qy     383 MQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAV-EDAASGMQEVSTATDEQAV 441
           :  |      |::  :|:    :||: :  :::||  : :  : :| : : : : |: |
Db     458 IVQEINAVSSQMNLNSQK----MERLSDMSKSVQETYEKMSSNLSSVVSDSNQSMDDYAK 513
```

```
Qy    442 STEEVAEMVDGVDDRAGEIAAALDDIADATD 472
          |  ::  |   | | |     |   :||::|
Db    514 SGHQIEVM---VSDFAEVEKVASKTLADSSD 541
```

RESULT 13
AAB46327
ID AAB46327 standard; Protein; 565 AA.
XX
AC AAB46327;
XX
DT 05-APR-2001 (first entry)
XX
DE H. pylori HPS024 protein.
XX
KW Microbial infection; antibacterial; Helicobacter pylori infection;
KW vaccine; screening.
XX
OS Helicobacter pylori.
XX
PN WO200073502-A2.
XX
PD 07-DEC-2000.
XX
PF 31-MAY-2000; 2000WO-EP05024.
XX
PR 31-MAY-1999; 99DE-1024965.
PR 17-JUN-1999; 99DE-1027740.
PR 21-JUL-1999; 99DE-1034029.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (CREA-) CREATOGEN GMBH.
XX
PI Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF;
XX
DR WPI; 2001-049948/06.
DR N-PSDB; AAF25604.
XX
PT Preparing an agent for diagnosis or control of microbial infection,
PT useful particularly against Helicobacter, based on identification of
PT essential genes in defective mutants -
XX
PS Claim 37; Page 289-291; 366pp; German.
XX
CC This invention describes a novel preparation of an agent (A) for
CC detection, prevention and/or treatment of microbial infection by:
CC (i) identifying essential genes (I) and corresponding polypeptides
CC (II); (ii) identifying compounds that are directed against (II) and
CC inactivate the microbe; (iii) testing these for suitability for use; and
CC (iv) formulating selected (A). Identifying essential genes (I) comprises
CC preparation of gene-deficient microorganisms by conditional antisense
CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
CC then determining viability and/or survival of the deficient organisms.
CC The products of the invention have antibacterial activity. (A) (which may
CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived
CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their
CC fragments or an inhibitor of (IIa)) are particularly used for diagnosis,
CC treatment or prevention of infection by Helicobacter pylori. Particularly
CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method
CC identifies essential genes, including those that have homologs in other
CC species, so identified (A) should have a broad spectrum of activity. Many
CC gene-deficient cells can be screened quickly, in an automated process,
CC and the identified genes can be used for screening without purification.
XX
SQ Sequence 565 AA;

```
  Query Match            12.1%;  Score 289;  DB 22;  Length 565;
  Best Local Similarity  27.2%;  Pred. No. 4.4e-13;
  Matches   90;  Conservative   69;  Mismatches  122;  Indels   50;  Gaps    9;

Qy    164 VDELVARFLPMLKLLTF-----DQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVE-- 216
          :||||      ||:  |     | :  :|         : |  |   ::   :   ||

Db    239 IDELV------LKINAFSRGDKDLRAKID---------VGDRNDEISQVGRGINLFVENA 283

Qy    217 -------APLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVAS--- 266
                  :|:|  || |   :       | :    | | |  ::||     ::||
Db    284 RLIMEEIKGISTLNKTSMDKLVQI------TQETQKSMKDSSTTLNSVKNKATDIASMMN 337

Qy    267 ----TADDVRRTSEDAEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVES 322
               :  :|:   : : | :: : |  |  : :|:     : ::: ||||   | ||:|
Db    338 ASIEQSQGLRKRLIETQGLVKESKDAIGDLFSQITESAHTEEELSSKVEQLSRNADDVKS 397

Qy    323 VTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQ 382
          :  :|:|||:|||:|||||:|||||||| | ||||||||: ||  :::    :  :
Db    398 ILDIINDIADQTNLLALNAAIEAARAGEHGRGFAVVADEVRNLAGRTQKSLAEINSTIMV 457

Qy    383 MQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEITDAV-EDAASGMQEVSTATDEQAV 441
          :  |      |::  :|:    :||: :  :::||  : :  : :| : : : : |: |
Db    458 IVQEINAVSSQMNLNSQK----MERLSDMSKSVQETYEKMSSNLSSVVSDSNQSMDDYAK 513

Qy    442 STEEVAEMVDGVDDRAGEIAAALDDIADATD 472
          |  ::  |   | | |     |   :||::|
Db    514 SGHQIEVM---VSDFAEVEKVASKTLADSSD 541
```

RESULT 14
AAW98797
ID AAW98797 standard; Protein; 630 AA.
XX
AC AAW98797;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 1278 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; AAX14516.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 8; Page 1738-1741; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 630 AA;

```
  Query Match            11.9%;  Score 284;  DB 19;  Length 630;
```

```
  Best Local Similarity   23.0%;  Pred. No. 1.2e-12;
  Matches  101;  Conservative   94;  Mismatches  145;  Indels  100;  Gaps   14;

Qy      70 DHLESYERTQDLFANSTKTVEQLKETQAE----YLLGLGRGEYDTEYAAQRARIGKIHDV 125
           ::|: :   ::  |   :| : | :|:::    ::: |   | |  |     ::| :  |
Db     206 EYLDPFSHKENFLA--VETFKMLGKTESKDNLNWMIAL-IIEKDKVY----EQVGSVRFV 258

Qy     126 LGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPML---------- 175
           : :   : : |     | |: |    :|: | |    ||   :: |  :|
Db     259 VIIASAIMVLALIIAITLLMRA----IVSSRLE----AVSSTLSHFFKLLNNQANSSGIK 310

Qy     176 -------------------------KLLTFDQQIAMDTY-------IDSYAQR------- 196
                                     |::  |:|   ||           ::| |
Db     311 LIEAKSNDELGRMQTAINKNILQTQKIMQEDRQAVQDTIKVVSDVKAGNFAVRITAEPAS 370

Qy     197 -----LHDEIDSRQE-LANAVATHVEAPLSSLEATS--------QDVAERTDTMRARTDD 242
               | | ::  : |  :| ||: :     |: |         |: : | : :
Db     371 PDLKELRDALNGIMDYLQESVGTHMPSIFKIFESYSGLDFRGRIQNASGRVELVTNALGQ 430

Qy     243 QVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEA 302
           :: :| : |      |   :::|: : :::   :: |  :      :  :    |: :|  :
Db     431 EIQKMLETS------SNFAKDLANDSANLKECVQNLEKASNSQHKSLMETSKTIENITTS 484

Qy     303 TDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEV 362
             ||::  | : |:  |::|:  :| |||:|||:|||||:|||||||| | |||||||||
Db     485 IQGVSSQSEAMIEQGQDIKSIVEIIRDIADQTNLLALNAAIEAARAGEHGRGFAVVADEV 544

Qy     363 KALAEESREQSTRVEE----LVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEI 418
           : ||| :::  : :|     ||: :   :|   :|: |        || :  ::| |: :
Db     545 RKLAERTQKSLSEIEANINILVQSISDTSESIKNQVKE--------VEEINASIEALRSV 596

Qy     419 TDAVEDAASGMQEVSTATDE 438
           |:     ||   |:|   |:
Db     597 TEGNLKIASDSLEISQEIDK 616


RESULT 15
AAW71553
ID   AAW71553 standard; Protein; 630 AA.
XX
AC   AAW71553;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Helicobacter polypeptide GHPO 1278.
XX
KW   GHPO 1278; infection; therapy; diagnosis; vaccine; gastritis;
KW   ulcer.
XX
OS   Helicobacter pylori.
XX
PN   WO9821225-A1.
XX
PD   22-MAY-1998.
XX
PF   14-NOV-1997;   97WO-US21353.
XX
PR   29-JUL-1997;   97US-0902615.
PR   14-NOV-1996;   96US-0749051.
PR   01-APR-1997;   97US-0831309.
PR   01-APR-1997;   97US-0833457.
PR   01-APR-1997;   97US-0834705.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Haas R,  Kleanthous H,  Meyer T,  Miller C;
PI   Odenbreit S,  Tomb J;
XX
DR   WPI; 1998-297855/26.
DR   N-PSDB; AAV52088.
XX
PT   Helicobacter polynucleotide and polypeptide sequences - useful to
PT   treat or prevent gastrointestinal infection
XX
PS   Claim 1; Page 308-310; 362pp; English.
XX
CC   This claimed Helicobacter pylori polypeptide, designated GHPO 1278,
CC   can be used in vaccination methods for preventing or treating
CC   Helicobacter infection.  85 Helicobacter polypeptides (see
CC   AAW71474-W71558) are claimed, as well as isolated polynucleotides
CC   (see AAV52009-93) that encode them.  The invention also provides:
CC   methods for producing these Helicobacter polypeptides in
CC   recombinant host systems, and related expression cassettes, vectors
CC   and transformed or transfected host cells; live vaccine vectors
CC   that contain the polynucleotides of the invention and which can be
CC   used to prevent or treat Helicobacter infection; therapeutic and/or
CC   prophylactic methods involving administration of polynucleotide
CC   molecules, polypeptides or monospecific antibodies; methods for
CC   detecting the presence of Helicobacter in samples using e.g.
CC   the polypeptides or monospecific antibodies; and methods for
CC   purifying the polypeptides by antibody-based affinity
CC   chromatography.
XX
SQ   Sequence   630 AA;

  Query Match             11.9%;  Score 284;  DB 19;  Length 630;
  Best Local Similarity   23.0%;  Pred. No. 1.2e-12;
  Matches  101;  Conservative   94;  Mismatches  145;  Indels  100;  Gaps   14;

Qy      70 DHLESYERTQDLFANSTKTVEQLKETQAE----YLLGLGRGEYDTEYAAQRARIGKIHDV 125
           ::|: :   ::  |   :| : | :|:::    ::: |   | |  |     ::| :  |
Db     206 EYLDPFSHKENFLA--VETFKMLGKTESKDNLNWMIAL-IIEKDKVY----EQVGSVRFV 258

Qy     126 LGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPML---------- 175
           : :   : : |     | |: |    :|: | |    ||   :: |  :|
Db     259 VIIASAIMVLALIIAITLLMRA----IVSSRLE----AVSSTLSHFFKLLNNQANSSGIK 310

Qy     176 -------------------------KLLTFDQQIAMDTY-------IDSYAQR------- 196
                                     |::  |:|   ||           ::| |
Db     311 LIEAKSNDELGRMQTAINKNILQTQKIMQEDRQAVQDTIKVVSDVKAGNFAVRITAEPAS 370

Qy     197 -----LHDEIDSRQE-LANAVATHVEAPLSSLEATS--------QDVAERTDTMRARTDD 242
               | | ::  : |  :| ||: :     |: |         |: : | : :
Db     371 PDLKELRDALNGIMDYLQESVGTHMPSIFKIFESYSGLDFRGRIQNASGRVELVTNALGQ 430

Qy     243 QVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQGEAAADDALATMTDIDEA 302
           :: :| : |      |   :::|: : :::   :: |  :      :  :    |: :|  :
Db     431 EIQKMLETS------SNFAKDLANDSANLKECVQNLEKASNSQHKSLMETSKTIENITTS 484

Qy     303 TDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFAVVADEV 362
             ||::  | : |:  |::|:  :| |||:|||:|||||:|||||||| | |||||||||
Db     485 IQGVSSQSEAMIEQGQDIKSIVEIIRDIADQTNLLALNAAIEAARAGEHGRGFAVVADEV 544

Qy     363 KALAEESREQSTRVEE----LVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMETLQEI 418
           : ||| :::  : :|     ||: :   :|   :|: |        || :  ::| |: :
Db     545 RKLAERTQKSLSEIEANINILVQSISDTSESIKNQVKE--------VEEINASIEALRSV 596

Qy     419 TDAVEDAASGMQEVSTATDE 438
           |:     ||   |:|   |:
Db     597 TEGNLKIASDSLEISQEIDK 616


Search completed: January  2, 2003, 12:41:39
Job time : 99 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 12:29:51 ; Search time 17 Seconds
(without alignments)
846.342 Million cell updates/sec

Title: US-09-455-978B-2
Perfect score: 2394
Sequence: 1 MSNDNDTLVTADVRNGIDGH.....ATDQQVRTVEEVRETVGKLS 489

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	205	8.6	955	2	US-08-428-414A-3	Sequence 3, Appli
2	202.5	8.5	955	1	US-08-006-676B-1	Sequence 1, Appli
3	202.5	8.5	955	1	US-08-282-845-2	Sequence 2, Appli
4	202.5	8.5	955	5	PCT-US94-00324-1	Sequence 1, Appli
5	194	8.1	1786	4	US-08-973-462-8	Sequence 8, Appli
6	188.5	7.9	885	2	US-08-533-306A-4	Sequence 4, Appli
7	188.5	7.9	885	2	US-08-742-923A-4	Sequence 4, Appli
8	187	7.8	630	4	US-08-973-462-9	Sequence 9, Appli
9	178	7.4	2101	1	US-08-466-390-4	Sequence 4, Appli
10	178	7.4	2101	1	US-08-470-950-4	Sequence 4, Appli
11	178	7.4	2101	1	US-08-467-781-4	Sequence 4, Appli
12	178	7.4	2101	1	US-08-195-487-4	Sequence 4, Appli
13	178	7.4	2101	2	US-08-483-924-4	Sequence 4, Appli
14	178	7.4	2101	4	US-09-452-294-1	Sequence 1, Appli
15	178	7.4	2101	5	PCT-US93-06160-4	Sequence 4, Appli
16	176	7.4	540	4	US-08-973-462-22	Sequence 22, Appl
17	173	7.2	1196	1	US-08-144-121-4	Sequence 4, Appli
18	173	7.2	1196	2	US-08-735-893-4	Sequence 4, Appli
19	170	7.1	816	2	US-08-533-306A-6	Sequence 6, Appli
20	170	7.1	816	2	US-08-742-923A-6	Sequence 6, Appli
21	164.5	6.9	534	4	US-09-103-664A-2	Sequence 2, Appli
22	160.5	6.7	1939	4	US-09-310-187A-1	Sequence 1, Appli
23	159	6.6	1018	1	US-08-072-610-2	Sequence 2, Appli
24	159	6.6	1018	2	US-08-719-822B-2	Sequence 2, Appli
25	159	6.6	1018	4	US-09-092-458-2	Sequence 2, Appli
26	156	6.5	1886	4	US-08-938-105-3	Sequence 3, Appli
27	156	6.5	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
28	154.5	6.5	1354	3	US-08-685-871-2	Sequence 2, Appli
29	152	6.3	576	2	US-08-533-306A-2	Sequence 2, Appli
30	152	6.3	576	2	US-08-742-923A-2	Sequence 2, Appli
31	151	6.3	1184	4	US-09-541-782-2	Sequence 2, Appli
32	151	6.3	1184	4	US-09-723-820-2	Sequence 2, Appli
33	151	6.3	2482	1	US-08-328-254-6	Sequence 6, Appli
34	151	6.3	3111	2	US-08-460-309-4	Sequence 4, Appli
35	151	6.3	3111	2	US-08-125-077-4	Sequence 4, Appli
36	150.5	6.3	566	2	US-08-533-669A-8	Sequence 8, Appli
37	150.5	6.3	566	2	US-08-511-872-2	Sequence 2, Appli
38	150.5	6.3	566	4	US-09-183-861-8	Sequence 8, Appli
39	150.5	6.3	566	4	US-09-022-765-8	Sequence 8, Appli
40	150	6.3	527	2	US-08-762-106-9	Sequence 9, Appli
41	150	6.3	527	4	US-09-320-774-9	Sequence 9, Appli
42	150	6.3	941	4	US-09-336-447A-9	Sequence 9, Appli
43	150	6.3	3248	1	US-08-353-700-1	Sequence 1, Appli
44	150	6.3	3248	5	PCT-US95-16216-1	Sequence 1, Appli
45	149	6.2	516	2	US-08-762-106-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-428-414A-3
; Sequence 3, Application US/08428414A
; Patent No. 5912166
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: LEISHMANIASIS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,414A
; FILING DATE: 21-APR-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Kadlecek, Ann T.
; REGISTRATION NUMBER: 39,244
; REFERENCE/DOCKET NUMBER: 210121.407
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANDBERRY
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 955 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-428-414A-3

Query Match 8.6%; Score 205; DB 2; Length 955;
Best Local Similarity 23.3%; Pred. No. 6e-08;
Matches 115; Conservative 80; Mismatches 198; Indels 100; Gaps 19;

```
Qy      13 VRNGIDGHALADRIGLDEAEIAWRLSFTGIDDTMAALAAEQPLFEATADALVT---DFY 69
           || :|    ::|  |              : |:| |  ||   |  ||| |
Db     530 VRRRLDAEIASEREKL--------------ESTVAQLEREQRREVALDALQTHQRKLQ 574

Qy      70 DHLESYERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARIGKIHDVLGLG 129
```

```
          : ||| |||    |   : ::|| | |:|      | :  ::    | |: :  |:  :
Db    575 EALESSERTA---AERDQLLQQLTELQSE------RTQL-SQVVTDRERLTR--DLQRI- 621

Qy    130 PDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVDELVARFLPMLKLLTFDQQIAMDTY 189
                :|  |  : || ||     :|  |      |     :|:| |          :
Db    622 ---------QYEYGETE-LARDVALCAAQEMEARYHAAVFHLQTLLELAT--------EW 663

Qy    190 IDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMAD 249
           |:  :|   | |        | |  ::|  |: :   :   ||  ::  :  :  :| |:
Db    664 EDALRERALAERD------EAAAAELDAAASTSQNARESACERLTSLEQQLRESEERAAE 717

Qy    250 VSREISSVSASVEEVASTADDVRRTSED--------AEALAQQGEAAADDALATMTDIDE 301
          :: :: : :|:        :: | | |          |  || | || |   ::    | :
Db    718 LASQLEATAAAKSSAEQDRENTRATLEQQLRESEARAAELASQLEATAAAKMSAEQDREN 777

Qy    302 ATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNA--SIEAARAGEAGEGFAVVA 359
          :   :   |||:: |          | |    ::|    |: ||
Db    778 TRATLEQQLRDSEERAAELASQ-------LESTTAAKMSAEQDRESTRA---------- 819

Qy    360 DEVKALAEESREQSTRVEELVEQMQAET--EETVDQLDE-----VNQRIGEGVERVEEAM 412
              | :: |:   |  ||  |::: |  : : :|  |      : |:: |  ||  |
Db    820 ----TLEQQLRDSEERAAELASQLESTTAAKMSAEQDRESTRATLEQQLRESEERAAELA 875

Qy    413 ETLQEITDAVEDAASGMQEVSTATDEQAV--STEEVAEMVDGVDDRAGEIAAALDDIAD- 469
            | | | : :|:    :| : || || :  | |  ||:   ::  |   ::|  |  :
Db    876 SQL-ESTTAAKMSAEQDRESTRATLEQQLRDSEERAAELASQLEATAAAKSSAEQDRENT 934

Qy    470 --ATDQQVRTVEE 480
            | :||:|  ||
Db    935 RAALEQQLRDSEE 947


RESULT 2
US-08-006-676B-1
; Sequence 1, Application US/08006676B
; Patent No. 5411865
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven
;    TITLE OF INVENTION:  Diagnosis of Leishmaniasis
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jeffrey B. Oster
;      STREET:  8339 SE 57th Street
;      CITY:  Mercer Island
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98040-4906
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatable
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORD for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/006,676B
;      FILING DATE:  15-JAN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oster, Jeffrey B.
;      REGISTRATION NUMBER:  32,585
;      REFERENCE/DOCKET NUMBER:  REED-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 232 7845
;      TELEFAX:  (206) 236 0205
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  955 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-006-676B-1

  Query Match             8.5%;  Score 202.5;  DB 1;  Length 955;
  Best Local Similarity   23.8%;  Pred. No. 9.4e-08;
  Matches  110;  Conservative   77;  Mismatches  190;  Indels   85;  Gaps   18;

Qy     44 DDTMAALAAEQPLFEATADALVT---DFYDHLESYERTQDLFANSTKTVEQLKETQAEYL 100
          : |:| |  ||   |  ||| |     : ||| |||    |  : ::|| | |:|
Db    546 ESTVAQLEREQRERERVALDALQTHQRKLQEALESSERTA---AERDQLLQQLTELQSE-- 600

Qy    101 LGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEA 160
              | :  ::    | |: :  |:  :           :|  |  : || ||     :|
Db    601 ----RTQL-SQVVTDRERLTR--DLQRI----------QYEYGETE-LARDVALCAAQEM 642

Qy    161 AAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLS 220
            |      |     :|:| |          : |:  :|   | |        | |  ::|  |
Db    643 EARYHAAVFHLQTLLELAT--------EWEDALRERALAERD------EAAAAELDAAAS 688

Qy    221 SLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED--- 277
          : :   :   ||  ::  :  :  :| |::: :: : :|:          :: | | |
Db    689 TSQNARESACERLTSLEQQLRESEERAAELASQLEATAAAKSSAEQDRENTRATLEQQLR 748

Qy    278 -----AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAE 332
               | || |• || |   ::   | :     :   :    ||||:: |        |
Db    749 ESEARAAELASQLEATAAAKMSAEQDRENTRATLEQQLRDSEERAAELASQ-------LE 801

Qy    333 QTNMLALNA--SIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAET--E 388
            |    ::|    |: ||                | :: |:   |  ||  |::: |  :
Db    802 STTAAKMSAEQDRESTRA--------------TLEQQLRDSEERAAELASQLESTTAAK 846

Qy    389 ETVDQLDE-----VNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAV-- 441
          : :|  |      : |:: |  ||  |     | | | | : :|·  :| : || || :
Db    847 MSAEQDRESTRATLEQQLRESEERAAELASQL-ESTTAAKMSAEQDRESTRATLEQQLRD 905

Qy    442 STEEVAEMVDGVDDRAGEIAAALDDIAD---ATDQQVRTVEE 480
            | |  ||:   ::  |   ::|  |  :    | :||:|  ||
Db    906 SEERAAELASQLEATAAAKSSAEQDRENTRAALEQQLRDSEE 947


RESULT 3
US-08-282-845-2
; Sequence 2, Application US/08282845
; Patent No. 5719263
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  A 230Kd Antigen Present in Leishmania
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple Macintosh Operating System 7.1
;      SOFTWARE:  Microsoft Word for Macintosh 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/282,845
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/006,676
;      FILING DATE:  JANUARY 15, 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia Anne
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  5004-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206)587-0430
;      TELEFAX:  (206)233-0644
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  955 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-282-845-2

  Query Match             8.5%;  Score 202.5;  DB 1;  Length 955;
  Best Local Similarity   23.8%;  Pred. No. 9.4e-08;
  Matches  110;  Conservative   77;  Mismatches  190;  Indels   85;  Gaps   18;

Qy      44 DDTMAALAAEQPLFEATADALVT---DFYDHLESYERTQDLFANSTKTVEQLKETQAEYL 100
           : |:| |  ||   |   ||| |      : ||| |||    |   : ::|| | |:|
Db     546 ESTVAQLEREQREREVALDALQTHQRKLQEALESSERTA---AERDQLLQQLTELQSE-- 600

Qy     101 LGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEA 160
           | :  ::    | |: :  |:  :          :|  |  : || ||     :|
Db     601 ----RTQL-SQVVTDRERLTR--DLQRI----------QYEYGETE-LARDVALCAAQEM 642

Qy     161 AAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLS 220
           |     |     :|:| |         : |:  :|   | |       | |  ::|  |
Db     643 EARYHAAVFHLQTLLELAT--------EWEDALRERALAERD------EAAAAELDAAAS 688

Qy     221 SLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED--- 277
           : :   :   ||  ::  :  :  :| |::: :: : :|:        :: | | |
Db     689 TSQNARESACERLTSLEQQLRESEERAAELASQLEATAAAKSSAEQDRENTRATLEQQLR 748

Qy     278 -----AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAE 332
                |  || | || |   ::   | :     :   :    |||:: |         |
Db     749 ESEARAAELASQLEATAAAKMSAEQDRENTRATLEQQLRDSEERAAELASQ-------LE 801

Qy     333 QTNMLALNA--SIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAET--E 388
           |    ::|    |: ||                | :: |:   |  ||  |::: |  :
Db     802 STTAAKMSAEQDRESTRA----------TLEQQLRDSEERAAELASQLESTTAAK 846

Qy     389 ETVDQLDE-----VNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAV-- 441
           : :|  |     : |:: |  ||  |    | | | | : :|   :| : || || :
Db     847 MSAEQDRESTRATLEQQLRESEERAAELASQL-ESTTAAKMSAEQDRESTRATLEQQLRD 905

Qy     442 STEEVAEMVDGVDDRAGEIAAALDDIAD---ATDQQVRTVEE 480
           | |  ||:   ::  |   ::|  |  :   | :||:|  ||
Db     906 SEERAAELASQLEATAAAKSSAEQDRENTRAALEQQLRDSEE 947


RESULT 4
PCT-US94-00324-1
; Sequence 1, Application PC/TUS9400324
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven
;    TITLE OF INVENTION:  Diagnosis of Leishmaniasis
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple System 7.1
;      SOFTWARE:  Microsoft Word, version 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/00324
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/006,676
;      FILING DATE:  15-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia Anne
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  5004-WO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 587-0430
;      TELEFAX:  (206) 233-0644
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  955 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US94-00324-1

  Query Match             8.5%;  Score 202.5;  DB 5;  Length 955;
  Best Local Similarity   23.8%;  Pred. No. 9.4e-08;
  Matches  110;  Conservative   77;  Mismatches  190;  Indels   85;  Gaps   18;

Qy      44 DDTMAALAAEQPLFEATADALVT---DFYDHLESYERTQDLFANSTKTVEQLKETQAEYL 100
           : |:| |  ||   |   ||| |      : ||| |||    |   : ::|| | |:|
Db     546 ESTVAQLEREQREREVALDALQTHQRKLQEALESSERTA---AERDQLLQQLTELQSE-- 600

Qy     101 LGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEA 160
           | :  ::    | |: :  |:  :          :|  |  : || ||     :|
Db     601 ----RTQL-SQVVTDRERLTR--DLQRI----------QYEYGETE-LARDVALCAAQEM 642

Qy     161 AAAVDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLS 220
           |     |     :|:| |         : |:  :|   | |       | |  ::|  |
Db     643 EARYHAAVFHLQTLLELAT--------EWEDALRERALAERD------EAAAAELDAAAS 688

Qy     221 SLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSED--- 277
           : :   :   ||  ::  :  :  :| |::: :: : :|:        :: | | |
Db     689 TSQNARESACERLTSLEQQLRESEERAAELASQLEATAAAKSSAEQDRENTRATLEQQLR 748

Qy     278 -----AEALAQQGEAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAE 332
                |  || | || |   ::   | :     :   :    |||:: |         |
Db     749 ESEARAAELASQLEATAAAKMSAEQDRENTRATLEQQLRDSEERAAELASQ-------LE 801

Qy     333 QTNMLALNA--SIEAARAGEAGEGFAVVADEVKALAEESREQSTRVEELVEQMQAET--E 388
           |    ::|    |: ||                | :: |:   |  ||  |::: |  :
Db     802 STTAAKMSAEQDRESTRA----------TLEQQLRDSEERAAELASQLESTTAAK 846

Qy     389 ETVDQLDE-----VNQRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAV-- 441
           : :|  |     : |:: |  ||  |    | | | | : :|   :| : || || :
Db     847 MSAEQDRESTRATLEQQLRESEERAAELASQL-ESTTAAKMSAEQDRESTRATLEQQLRD 905

Qy     442 STEEVAEMVDGVDDRAGEIAAALDDIAD---ATDQQVRTVEE 480
           | |  ||:   ::  |   ::|  |  :   | :||:|  ||
Db     906 SEERAAELASQLEATAAAKSSAEQDRENTRAALEQQLRDSEE 947


RESULT 5
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8
```

```
  Query Match             8.1%;  Score 194;  DB 4;  Length 1786;
  Best Local Similarity   24.5%;  Pred. No. 1e-06;
  Matches  100;  Conservative   85;  Mismatches  139;  Indels   84;  Gaps   22;

Qy    146 DALADDV---VADRGEEAAAA-VDELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEI 201
          :::|::|   ||:  ||: |  |:|:||  :  :   | :: :|  : ::| |  : : :
Db    565 ESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVA-PSVVESVAPSVEESV 623

Qy    202 DSRQELANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREI------S 255
          :   |   :|| :||   |  |   : ||| : : | | :::  :|    ||
Db    624 EENVE--ESVAENVEE--SVAENVEESVAENVEEIVAPTVEEI--VAPTVEEIVAPSVVE 677

Qy    256 SVSASVEEVA-----STADDVRRT-SEDAEALAQQGEAAADDALATMTD-IDEATDGVT 307
          ||: ||||       | |::|  : :|: |      |: |::   :: : ::|    |
Db    678 SVAPSVEESVEENVEESVAENVEESVAENVE-----ESVAENVEESVAENVEEI---VA 728

Qy    308 AGVEQLGERAAD-------VESVTGVIDDIAEQTNMLALNASIEAARAGEAGEGFA---- 356
            ||::     :       ||||   :::  |:    ::  ::| : |    |  |
Db    729 PTVEEIVAPTVEEIVAPSVVESVAPSVEESVEENVEESVAENVEESVAENVEESVAENVE 788

Qy    357 -VVADEVKALAEESREQST--RVEELVEQMQA--------------ETEETVD----QL 394
           ||  |: :   | |:|    ||| | :  |               ||||  |    ::
Db    789 ESVAPTVEEIVAPSVEESVAPSVEESVAENVATNLSDNLLSNLLGGIETEEIKDSILNEI 848

Qy    395 DEVNQRIGEGV-ERVEEA----------METLQEIT---DAVEDAASGMQE-VSTATDE 438
          :|| : :   :| | |||           :| :|| |   | :|:    : | | :|  |
Db    849 EEVKENVVTTILENVEETTAESVTTFSNILEEIQENTITNDTIEEKLEELHENVLSAALE 908

Qy    439 QAVSTEEVAEMVDGVDDRAGEIAAAL-DDIADATDQQVRTVEEVRETV 485
            | ||  |:| | :::   |:|  | : :  | ::   |: |: | :
Db    909 NTQSEEEKKEVIDVIEEVKEEVATTLIETVEQAEEKSANTITEIFENL 956


RESULT 6
US-08-533-306A-4
; Sequence 4, Application US/08533306A
; Patent No. 5837457
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,306A
;      FILING DATE:  September 25, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869COB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-533-306A-4

  Query Match             7.9%;  Score 188.5;  DB 2;  Length 885;
  Best Local Similarity   21.7%;  Pred. No. 1.1e-06;
  Matches  113;  Conservative   98;  Mismatches  178;  Indels  131;  Gaps   23;

Qy     44 DDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEY---L 100
          :| :|  | |:       |     :| |   |:    :  ||  |  : |::  |:    |
Db    134 EDALAQQAFEE------ARRRTREFEDRDRSHREEMEAKANLDKNKQTLEKENADLAGEL 187

Qy    101 LGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVA--DRGE 158
            ||: : : |:  ::                                  |:|   :: :    ||
Db    188 RVLGQAKQEVEHKKKK-------------------------LEAQVQELQSKCSDGE 219

Qy    159 EAAAAVDELVARFLPMLKLLT------FDQQIAMDTYIDSYAQRLHD-----EIDSRQEL 207
          | | ::: | :    :: :|       : | :   : | : :| |     : ::||:|
Db    220 RARAELNDKVHKLQNEVESVTGMLNEAEGKAIKLAKDVASLSSQLQDTQELLQEETRQKL 279

Qy    208 ANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVS-------AS 260
              :|   |  ||       |  :::: : |::::   :: | ||:::
Db    280 ------NVSTKLRQLE-------EERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKK 326

Qy    261 VEEVASTADDV----RRTSEDAEALAQQGE--AAADDAL-ATMTDIDEATDGVTAGVEQL 313
          ::: ||| : :    :|  :: | | || |  ||| | |  |   : :  | :   ::
Db    327 LQDFASTVEALEEGKKRFQKEIENLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQ 386

Qy    314 GERAADVESVTGVIDD-IAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALA-----E 367
          :  :::|      |  :||: |: :  |        || |       | |||:     |
Db    387 RQLVSNLEKKQRKFDQLLAEEKNISSKYADERDRAEAEARE-----KETKALSLARALE 440

Qy    368 ESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMET--------LQEIT 419
          |: |    :|   : ::|| |: |   |:| : : | :|: : |:||        |:|:
Db    441 EALEAKEELERTNKMLKAEMEDLVSSKDDVGKNVHE-LEKSKRALETQMEEMKTQLEELE 499

Qy    420 D---AVEDA----ASGMQEVS-------TATDEQAVST---------EEVAEMVDGVDDR 456
          |   | |||       || :         | |||             |   |: |  ::|
Db    500 DELQASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERNER 559

Qy    457 AGEIAAA--------LDDIADATDQQVRTVEEVRETVGKL 488
          |  :|||         | |:    |  ::  ||  : : ||
Db    560 A--LAAAAKKKLEGDLKDLELQADSAIKGREEAIKQLRKL 597


RESULT 7
US-08-742-923A-4
; Sequence 4, Application US/08742923A
; Patent No. 5869611
;  GENERAL INFORMATION:
;    APPLICANT:  Liu, Pu
;    APPLICANT:  Collins, Francis S.
;    APPLICANT:  Siciliano, Michael J.
;    APPLICANT:  Claxton, David
;    TITLE OF INVENTION:  Markers for Detection of Chromosome 16
;    TITLE OF INVENTION:  Rearrangements
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  USA
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/742,923A
;      FILING DATE:  No. 5869611ember 1, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-00869DVC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 641-1600
;      TELEFAX:  (810) 641-0270
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  885 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-742-923A-4

  Query Match             7.9%;  Score 188.5;  DB 2;  Length 885;
  Best Local Similarity   21.7%;  Pred. No. 1.1e-06;
  Matches  113;  Conservative   98;  Mismatches  178;  Indels  131;  Gaps   23;

Qy      44 DDTMAALAAEQPLFEATADALVTDFYDHLESYERTQDLFANSTKTVEQLKETQAEY---L 100
           :| :|  | |:       |     :| |   |:    :  ||  |  : |::  |:    |
Db     134 EDALAQQAFEE------ARRRTREFEDRDRSHREEMEAKANLDKNKQTLEKENADLAGEL 187

Qy     101 LGLGRGEYDTEYAAQRARIGKIHDVLGLGPDVYLGAYTRYYTGLLDALADDVVA--DRGE 158
             ||: : : |:  ::                          |:|   :: :    ||
Db     188 RVLGQAKQEVEHKKKK--------------------------LEAQVQELQSKCSDGE 219

Qy     159 EAAAAVDELVARFLPMLKLLT------FDQQIAMDTYIDSYAQRLHD-----EIDSRQEL 207
             | | ::: | :    :: :|        : | :   : | : :| |     : ::||:|
Db     220 RARAELNDKVHKLQNEVESVTGMLNEAEGKAIKLAKDVASLSSQLQDTQELLQEETRQKL 279

Qy     208 ANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVS-------AS 260
               :|   |  ||       |  :::: : |::::   :: | ||:::
Db     280 ------NVSTKLRQLE-------EERNSLQDQLDEEMEAKQNLERHISTLNIQLSDSKKK 326

Qy     261 VEEVASTADDV----RRTSEDAEALAQQGE--AAADDAL-ATMTDIDEATDGVTAGVEQL 313
           ::: ||| : :    :|  :: | | || |  ||| | |  |   : :  | :   ::
Db     327 LQDFASTVEALEEGKKRFQKEIENLTQQYEEKAAAYDKLEKTKNRLQQELDDLVVDLDNQ 386

Qy     314 GERAADVESVTGVIDD-IAEQTNMLALNASIEAARAGEAGEGFAVVADEVKALA-----E 367
             :  :::|      |  :||: |: :  |         || |       | |||:     |
Db     387 RQLVSNLEKKQRKFDQLLAEEKNISSKYADERDRAEAEARE------KETKALSLARALE 440

Qy     368 ESREQSTRVEELVEQMQAETEETVDQLDEVNQRIGEGVERVEEAMET--------LQEIT 419
           |: |    :|  : ::|| |: |   |:| : : | :|: : |:||         |:|:
Db     441 EALEAKEELERTNKMLKAEMEDLVSSKDDVGKNVHE-LEKSKRALETQMEEMKTQLEELE 499

Qy     420 D---AVEDA----ASGMQEVS-------TATDEQAVST---------EEVAEMVDGVDDR 456
           |   | |||       || :         | |||              |   |: |  ::|
Db     500 DELQASEDAKLRLEVNMQALKGQFERDLQARDEQNEEKRRQLQRQLHEYETELEDERNER 559

Qy     457 AGEIAAA--------LDDIADATDQQVRTVEEVRETVGKL 488
           |  :|||        | |:    |  ::  ||  : : ||
Db     560 A--LAAAAKKKLEGDLKDLELQADSAIKGREEAIKQLRKL 597


RESULT 8
US-08-973-462-9
; Sequence 9, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;  LENGTH: 630
;  TYPE: PRT
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-9

  Query Match             7.8%;  Score 187;  DB 4;  Length 630;
  Best Local Similarity   21.5%;  Pred. No. 8.8e-07;
  Matches   85;  Conservative   87;  Mismatches  138;  Indels   86;  Gaps   16;

Qy     150 DDVVADRGEEAAA-AVDELVARFLPMLK--LLTFDQQIAMDTYIDSYAQRLHDEIDSRQE 206
           :::||   ||: |  |:|:||   | ::  :   ::| : :  :| |  : : :   :
Db     263 EEIVAPTVEESVAPTVEEIVA---PTVEESVAPTVEEIVVPSVEESVAPSVEESV--AEN 317

Qy     207 LANAVATHVEAPLSSLEATSQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVA- 265
           :  :|| :||   |  |   : |||  :   |  :::            |: ||||:
Db     318 VEESVAENVEE--SVAENVEESVAENVEESVAENVEEI-----------VAPSVEEIVA 363

Qy     266 -----STADDVRRTSEDAEALAQ-----QGEAAADDALATMTDIDEATDGVTAGVEQLGE 315
                | |::|     |   |:      : |  |  |  : :: |  : ||  :|:: |
Db     364 PTVEESVAENVATNLSD-NLLSNLLGGIETEEIKDSILNEIEEVKE--NVVTTILEKVEE 420

Qy     316 RAAD-VESVTGVIDDIAEQT------------NMLALNASIEAARAGEAGEGFAVVADE 361
             |: | : : ::::| | |             :   |:|::|  :: |  :    | :|
Db     421 TTAESVTTFSNILEEIQENTITNDTIEEKLEELHENVLSAALENTQSEEEKKEVIDVIEE 480

Qy     362 VKALAEESREQSTRVEELVEQMQAETEETVDQL--------DEVNQRIGEGVERVEEAM- 412
           ||       | :| : | ||| : |:| |: ::          | |::: | :|:: | :
Db     481 VK------EEVATTLIETVEQAEEESESTITEIFENLEENAVESNEKVAENLEKLNETVF 534

Qy     413 -ETLQEITDAVE-------------------DAASGMQEVSTATDEQAVSTEEVAEMVD 451
            | :: : ||                    |   |:||        ::: |  | :  :
Db     535 NTVLDKVEETVEISGESLENNEMDKAFFSEIFDNVKGIQENLLTGMFRSIETSIVIQSEE 594

Qy     452 GVDDRAGEIAAALDDIADATDQQVRTVEEVRETVGK 487
             ||     ::: ||:| :  :  :  :| :  | |:
Db     595 KVDLNENVVSSILDNIENMKEGLLNKLENISSTEGE 630


RESULT 9
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,390
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
```

```
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-466-390-4

  Query Match             7.4%;  Score 178;  DB 1;  Length 2101;
  Best Local Similarity   18.9%;  Pred. No. 2.2e-05;
  Matches   86;  Conservative   95;  Mismatches  192;  Indels   82;  Gaps   14;

Qy      67 DFYDHLES-YERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARI--GKIH 123
           :|  ||:   :   :|    :|  ::  | ||:    |     | :   ::  |  ||:
Db     329 EFASHLQQLQDALNELTEEHSKATQEWLEKQAQLEKELSAALQDKKCLEEKNEILQGKLS 388

Qy     124 ------------------DVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVD 165
                            :|||                  | |  : :       ||
Db     389 QLEEHLSQLQDNPPQEKGEVLG------------------DVLQLETLKQEAATLAANNT 430

Qy     166 ELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEAT 225
           :| || : ||:     |:  :        |:| | | : :|:|:: : | ::: :|:|
Db     431 QLQAR-VEMLETERGQQEAKL------LAERGHFE-EEKQQLSSLI-TDLQSSISNLSQA 481

Qy     226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
            ::: : :   .||   ||   | :: |:::::|::::             ::   | ||
Db     482 KEELEQASQAHGARLTAQV---ASLTSELTTLNATIQQ----------QDQELAGLKQQA 528

Qy     286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
           :         |:  ::|: |:   ||||        |     : ::||:       : : :
Db     529 KEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQ---QLKEVAEKQEATRQDHAQQL 585

Qy     346 ARAGEAGEGFAVVAD-EVKALAEESREQSTRVEELVEQMQAETE------ETVDQLDEVN 398
           | | |  |       |  :| |    :|:: ::| | :|:|   |        :| |
Db     586 ATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREK 645

Qy     399 QRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR-- 456
           :   || ::  :|| ::    :   : ::    :  ::|   | ||:  | : ::
Db     646 AELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQQKATEKERVAQEKDQLQEQLQ 705

Qy     457 ---------AGEIAAALDDIADATDQQVRTVEEVR 482
                   | :       ||| ::| | : |::
Db     706 ALKESLKVTKGSLEEEKRRAADALEEQQRCISELK 740


RESULT 10
US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5698439
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,950
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-470-950-4

  Query Match             7.4%;  Score 178;  DB 1;  Length 2101;
  Best Local Similarity   18.9%;  Pred. No. 2.2e-05;
  Matches   86;  Conservative   95;  Mismatches  192;  Indels   82;  Gaps   14;

Qy      67 DFYDHLES-YERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARI--GKIH 123
           :|  ||:   :   :|    :|  ::  | ||:    |     | :   ::  |  ||:
Db     329 EFASHLQQLQDALNELTEEHSKATQEWLEKQAQLEKELSAALQDKKCLEEKNEILQGKLS 388

Qy     124 ------------------DVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVD 165
                            :|||                  | |  : :       ||
Db     389 QLEEHLSQLQDNPPQEKGEVLG------------------DVLQLETLKQEAATLAANNT 430

Qy     166 ELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEAT 225
           :| || : ||:     |:  :        |:| | | : :|:|:: : | ::: :|:|
Db     431 QLQAR-VEMLETERGQQEAKL------LAERGHFE-EEKQQLSSLI-TDLQSSISNLSQA 481

Qy     226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
            ::: : :   .||   ||   | :: |:::::|::::             ::   | ||
Db     482 KEELEQASQAHGARLTAQV---ASLTSELTTLNATIQQ----------QDQELAGLKQQA 528

Qy     286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
           :         |:  ::|: |:   ||||        |     : ::||:       : : :
Db     529 KEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQ---QLKEVAEKQEATRQDHAQQL 585

Qy     346 ARAGEAGEGFAVVAD-EVKALAEESREQSTRVEELVEQMQAETE------ETVDQLDEVN 398
           | | |  |       |  :| |    :|:: ::| | :|:|   |        :| |
Db     586 ATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREK 645

Qy     399 QRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR-- 456
           :   || ::  :|| ::    :   : ::    :  ::|   | ||:  | : ::
Db     646 AELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQQKATEKERVAQEKDQLQEQLQ 705

Qy     457 ---------AGEIAAALDDIADATDQQVRTVEEVR 482
                   | :       ||| ::| | : |::
Db     706 ALKESLKVTKGSLEEEKRRAADALEEQQRCISELK 740


RESULT 11
US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/467,781
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-467-781-4

  Query Match             7.4%;  Score 178;  DB 1;  Length 2101;
  Best Local Similarity   18.9%;  Pred. No. 2.2e-05;
  Matches   86;  Conservative   95;  Mismatches  192;  Indels   82;  Gaps   14;

Qy      67 DFYDHLES-YERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARI--GKIH 123
           :|  ||:   :   :|    :|  ::  | ||:    |     | :   ::  |  ||:
Db     329 EFASHLQQLQDALNELTEEHSKATQEWLEKQAQLEKELSAALQDKKCLEEKNEILQGKLS 388

Qy     124 ------------------DVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVD 165
                            :|||                 | |  : :       ||
Db     389 QLEEHLSQLQDNPPQEKGEVLG------------------DVLQLETLKQEAATLAANNT 430

Qy     166 ELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEAT 225
           :| || : ||:     |:  :       |:| | | : :|:|:: : | ::: :|:|
Db     431 QLQAR-VEMLETERGQQEAKL------LAERGHFE-EEKQQLSSLI-TDLQSSISNLSQA 481

Qy     226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
           ::: : :    ||   ||   | :: |:::::|:::            ::   | ||
Db     482 KEELEQASQAHGARLTAQV---ASLTSELTTLNATIQQ----------QDQELAGLKQQA 528

Qy     286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
           :       |:   ::|: |:   ||||      |     : ::||:      : : :
Db     529 KEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQ---QLKEVAEKQEATRQDHAQQL 585

Qy     346 ARAGEAGEGFAVVAD-EVKALAEESREQSTRVEELVEQMQAETE------ETVDQLDEVN 398
           | | |  |       |  :| |    :|:: ::| | :|:|   |       :| |
Db     586 ATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREK 645

Qy     399 QRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR-- 456
             :   || ::  :|| ::     :   : ::    :  ::|   | ||:  | : ::
Db     646 AELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQQKATEKERVAQEKDQLQEQLQ 705

Qy     457 ---------AGEIAAALDDIADATDQQVRTVEEVR 482
                   | :       ||| ::| | : |::
Db     706 ALKESLKVTKGSLEEEKRRAADALEEQQRCISELK 740


RESULT 12
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/195,487
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/901,701
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-195-487-4

  Query Match             7.4%;  Score 178;  DB 1;  Length 2101;
  Best Local Similarity   18.9%;  Pred. No. 2.2e-05;
  Matches   86;  Conservative   95;  Mismatches  192;  Indels   82;  Gaps   14;

Qy      67 DFYDHLES-YERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARI--GKIH 123
           :|  ||:   :   :|    :|  ::  | ||:    |     | :   ::  |  ||:
Db     329 EFASHLQQLQDALNELTEEHSKATQEWLEKQAQLEKELSAALQDKKCLEEKNEILQGKLS 388

Qy     124 ------------------DVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVD 165
                            :|||                 | |  : :       ||
Db     389 QLEEHLSQLQDNPPQEKGEVLG------------------DVLQLETLKQEAATLAANNT 430

Qy     166 ELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEAT 225
           :| || : ||:     |:  :       |:| | | : :|:|:: : | ::: :|:|
Db     431 QLQAR-VEMLETERGQQEAKL------LAERGHFE-EEKQQLSSLI-TDLQSSISNLSQA 481

Qy     226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
           ::: : :    ||   ||   | :: |:::::|:::            ::   | ||
Db     482 KEELEQASQAHGARLTAQV---ASLTSELTTLNATIQQ----------QDQELAGLKQQA 528

Qy     286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
           :       |:   ::|: |:   ||||      |     : ::||:      : : :
Db     529 KEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQ---QLKEVAEKQEATRQDHAQQL 585

Qy     346 ARAGEAGEGFAVVAD-EVKALAEESREQSTRVEELVEQMQAETE------ETVDQLDEVN 398
           | | |  |       |  :| |    :|:: ::| | :|:|   |       :| |
Db     586 ATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREK 645

Qy     399 QRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR-- 456
             :   || ::  :|| ::     :   : ::    :  ::|   | ||:  | : ::
Db     646 AELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQQKATEKERVAQEKDQLQEQLQ 705

Qy     457 ---------AGEIAAALDDIADATDQQVRTVEEVR 482
                   | :       ||| ::| | : |::
Db     706 ALKESLKVTKGSLEEEKRRAADALEEQQRCISELK 740


RESULT 13
US-08-483-924-4
; Sequence 4, Application US/08483924
; Patent No. 5882876
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,924
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-483-924-4

  Query Match             7.4%;  Score 178;  DB 2;  Length 2101;
  Best Local Similarity   18.9%;  Pred. No. 2.2e-05;
  Matches   86;  Conservative   95;  Mismatches  192;  Indels   82;  Gaps   14;

Qy      67 DFYDHLES-YERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARI--GKIH 123
           :|  ||:   :  :|    :|  ::  | ||:    |     | :   ::  |  ||:
Db     329 EFASHLQQLQDALNELTEEHSKATQEWLEKQAQLEKELSAALQDKKCLEEKNEILQGKLS 388

Qy     124 ------------------DVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVD 165
                            :|||                | |  : :        ||
Db     389 QLEEHLSQLQDNPPQEKGEVLG-----------------DVLQLETLKQEAATLAANNT 430

Qy     166 ELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEAT 225
           :| || : ||:     |:  :       |:| | | : :|:|:: : | ::: :|:|
Db     431 QLQAR-VEMLETERGQQEAKL------LAERGHFE-EEKQQLSSLI-TDLQSSISNLSQA 481

Qy     226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
           ::: : :    ||   ||   | :: |:::::|:::            ::   | ||
Db     482 KEELEQASQAHGARLTAQV---ASLTSELTTLNATIQQ----------QDQELAGLKQQA 528

Qy     286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
           :        |:   ::|: |:   ||||       |     : ::||:       : : :
Db     529 KEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQ---QLKEVAEKQEATRQDHAQQL 585

Qy     346 ARAGEAGEGFAVVAD-EVKALAEESREQSTRVEELVEQMQAETE------ETVDQLDEVN 398
           | | |  |      |  :| |    :|:: ::| | :|:|   |       :| |
Db     586 ATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREK 645

Qy     399 QRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR-- 456
             :   || ::  :|| ::     :   : ::     :  ::|   | ||:  | : ::
Db     646 AELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQQKATEKERVAQEKDQLQEQLQ 705

Qy     457 ---------AGEIAAALDDIADATDQQVRTVEEVR 482
                   | :       ||| ::| | : |::
Db     706 ALKESLKVTKGSLEEEKRRAADALEEQQRCISELK 740


RESULT 14
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
;  APPLICANT:  Lelievre, Sophie
;  APPLICANT:  Bissell, Mina
;  TITLE OF INVENTION:  UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
;  TITLE OF INVENTION:  THERAPH AND DETECTION OF PROLIFERATIVE AND
;  TITLE OF INVENTION:  DIFFERENTIATION DISORDERS
;  FILE REFERENCE:  IB-1454- Sequence Submittal
; Patent No. 6287790
;  CURRENT APPLICATION NUMBER: US/09/452,294
;  CURRENT FILING DATE:  1999-11-30
;  PRIOR APPLICATION NUMBER:  60/110,420
;  PRIOR FILING DATE:  1998-11-30
;  NUMBER OF SEQ ID NOS:  1
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 2101
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-452-294-1

  Query Match             7.4%;  Score 178;  DB 4;  Length 2101;
  Best Local Similarity   18.9%;  Pred. No. 2.2e-05;
  Matches   86;  Conservative   95;  Mismatches  192;  Indels   82;  Gaps   14;

Qy      67 DFYDHLES-YERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARI--GKIH 123
           :|  ||:   :  :|    :|  ::  | ||:    |     | :   ::  |  ||:
Db     329 EFASHLQQLQDALNELTEEHSKATQEWLEKQAQLEKELSAALQDKKCLEEKNEILQGKLS 388

Qy     124 ------------------DVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVD 165
                            :|||                | |  : :        ||
Db     389 QLEEHLSQLQDNPPQEKGEVLG-----------------DVLQLETLKQEAATLAANNT 430

Qy     166 ELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEAT 225
           :| || : ||:     |:  :       |:| | | : :|:|:: : | ::: :|:|
Db     431 QLQAR-VEMLETERGQQEAKL------LAERGHFE-EEKQQLSSLI-TDLQSSISNLSQA 481

Qy     226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
           ::: : :    ||   ||   | :: |:::::|:::            ::   | ||
Db     482 KEELEQASQAHGARLTAQV---ASLTSELTTLNATIQQ----------QDQELAGLKQQA 528

Qy     286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
           :        |:   ::|: |:   ||||       |     : ::||:       : : :
Db     529 KEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQ---QLKEVAEKQEATRQDHAQQL 585

Qy     346 ARAGEAGEGFAVVAD-EVKALAEESREQSTRVEELVEQMQAETE------ETVDQLDEVN 398
           | | |  |      |  :| |    :|:: ::| | :|:|   |       :| |
Db     586 ATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREK 645

Qy     399 QRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR-- 456
             :   || ::  :|| ::     :   : ::     :  ::|   | ||:  | : ::
Db     646 AELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQQKATEKERVAQEKDQLQEQLQ 705

Qy     457 ---------AGEIAAALDDIADATDQQVRTVEEVR 482
                   | :       ||| ::| | : |::
Db     706 ALKESLKVTKGSLEEEKRRAADALEEQQRCISELK 740


RESULT 15
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06160
;      FILING DATE:  19930621
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US93-06160-4

  Query Match             7.4%;  Score 178;  DB 5;  Length 2101;
  Best Local Similarity   18.9%;  Pred. No. 2.2e-05;
  Matches   86;  Conservative   95;  Mismatches  192;  Indels   82;  Gaps   14;

Qy      67 DFYDHLES-YERTQDLFANSTKTVEQLKETQAEYLLGLGRGEYDTEYAAQRARI--GKIH 123
           :|  ||:   :  :|   :|  ::  | ||:    |     | :   ::  |  ||:
Db     329 EFASHLQQLQDALNELTEEHSKATQEWLEKQAQLEKELSAALQDKKCLEEKNEILQGKLS 388

Qy     124 ------------------DVLGLGPDVYLGAYTRYYTGLLDALADDVVADRGEEAAAAVD 165
                             :|||                  | |  : :       ||
Db     389 QLEEHLSQLQDNPPQEKGEVLG-----------------DVLQLETLKQEAATLAANNT 430

Qy     166 ELVARFLPMLKLLTFDQQIAMDTYIDSYAQRLHDEIDSRQELANAVATHVEAPLSSLEAT 225
           :| || : ||:     |:  :       |:| | | : :|:|:: : | ::: :|:|
Db     431 QLQAR-VEMLETERGQQEAKL------LAERGHFE-EEKQQLSSLI-TDLQSSISNLSQA 481

Qy     226 SQDVAERTDTMRARTDDQVDRMADVSREISSVSASVEEVASTADDVRRTSEDAEALAQQG 285
            ::: : :   ||  ||   | :: |:::::|:::           ::   | ||
Db     482 KEELEQASQAHGARLTAQV---ASLTSELTTLNATIQQ----------QDQELAGLKQQA 528

Qy     286 EAAADDALATMTDIDEATDGVTAGVEQLGERAADVESVTGVIDDIAEQTNMLALNASIEA 345
           :       |:   ::|: |:   ||||    .  |     : ::||:      : : :
Db     529 KEKQAQLAQTLQQQEQASQGLRHQVEQLSSSLKQKEQ---QLKEVAEKQEATRQDHAQQL 585
                           |
Qy     346 ARAGEAGEGFAVVAD-EVKALAEESREQSTRVEELVEQMQAETE------ETVDQLDEVN 398
           | | |  |       |  :| |    :|:: ::| | :|:|   |       :| |
Db     586 ATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQAQREK 645

Qy     399 QRIGEGVERVEEAMETLQEITDAVEDAASGMQEVSTATDEQAVSTEEVAEMVDGVDDR-- 456
             :   || ::  :|| ::     :   : ::    :  ::|   | ||:  | : ::
Db     646 AELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQQKATEKERVAQEKDQLQEQLQ 705

Qy     457 ---------AGEIAAALDDIADATDQQVRTVEEVR 482
                   | :       ||| ::| | : |::
Db     706 ALKESLKVTKGSLEEEKRRAADALEEQQRCISELK 740


Search completed: January  2, 2003, 12:36:40
Job time : 23 secs
```